An investigation of genetic heterogeneity in a biological sentinel species (*Lumbricus rubellus*)

Robert Kevin Donnelly

A submission presented in partial fulfilment of the

requirements of the University of Glamorgan/Prifysgol Morgannwg

for the degree of Doctor of Philosophy

This research programme was carried out

in collaboration with Cardiff University

September 2011

Summary

Studies have indicated the existence of several possible sources of genetic heterogeneity within the sentinel species *Lumbricus rubellus* that could compromise its suitability for ecotoxicological assessment. The species appears to consist of two genetically divergent cryptic lineages and has been demonstrated to display genetic adaptation towards contaminants within some populations.

In order to investigate the cryptic lineages of *L. rubellus* further a combined morphological and DNA barcoding approach was undertaken, with both the mitochondrial COI gene being sequenced and external morphological characters being assessed. Combined barcoding and morphological analysis confirmed the existence of the previously described genetic lineages of *L. rubellus* and highlighted a potential lineage-specific morphological trait. The effectiveness of this trait in field identification of the two lineages was tested in a blind trial. This indicated that the trait may be particularly effective in successfully identifying the individuals of one of these lineages.

The two cryptic lineages were also analysed in a second study featuring cross-amplifying microsatellite loci. Both the sequencing and fragment analysis of these microsatellite loci strongly supported the existence of a high degree of reproductive isolation between the two lineages.

Finally microsatellite markers were applied to test the hypothesis of genetic adaptation within *L. rubellus* populations located along an aerially-deposited nickel contamination gradient. No support was obtained for the existence of genetic adaptation within these populations. This could indicate that the most heavily contaminated sites represent demographic 'sinks' into which individuals immigrate from adjacent less-contaminated regions.

The general implications of these studies are that the two lineages of *L. rubellus* should be treated as separate species in future ecotoxicological trials given their high degree of genetic differentiation. An implication of the nickel study is that *L. rubellus* may display a greater capacity for tolerating toxic metals through phenotypic plasticity than previous studies have indicated.

Acknowledgements

I wish to thank the following people for their help over the course of my PhD. Firstly I would like to thank my supervisor Mike Bruford for both allowing me to use his laboratory and for his patience and unfaltering enthusiasm over the course of this study. This was much appreciated, particularly during those times when my own enthusiasm seemed to be fading. I must also thank John Morgan for teaching me a great deal about both earthworms and ecotoxicology and for collecting many hundreds of earthworms. Without John's sixth sense for finding worms this project would have been a lot more difficult. I must also thank George Harper for her supervision and help at the start of the PhD, and for developing some of the microsatellites that I used in the project. Finally I would like to thank Denis Murphy for his help as my second supervisor.

Among the people at the University of Glamorgan who have helped me over the course of my PhD I would like to particularly thank Paul Hallas for being a great flat-mate, lab-mate and office-mate. I would also like to thank the following people at Glamorgan- Amie Noone, Mark Partridge, Roy Wiles, Martin Powell, Sylvia Davies, Alvine Jones, Ed Morgan, Chris Lee, John Morton and anyone else I might have missed.

Among the people at Cardiff University who have helped me over the course of my PhD I would like to particularly thank Gabi Juma who processed many of the earthworm samples analysed in the project, genotyped the worms in the nickel study to determine their lineage and determined the metal content of the soils. I am also grateful to Gabi for teaching me many molecular techniques at the start of the study. I must also thank Jane Andre and Andy King for their help and advice with some of the molecular work, Pete Kille for allowing me to use his laboratory and his advice on genetic barcoding and Pierfrancesco Sechi for genotyping the worms from Mid-Wales and England. I would also like to thank the following people at Cardiff- Adam, Alex, Asniza, Ben, Bill, Catherine, Coralie, Dave B, Dave S, Eddie, Ellie, Geoff, Isa-Rita, James P, James R, Jeff, Jo B, Jo C, Jo L, Jo S, Joan, Leila, Loys, Luca, Lucija, Maf B, Maf C, Milena, Mireille, Niall, Pablo, Pati, Renata, Rhys, Rui, Samrat, Sigrid, Steve, Tania, Tim, Vicky, Viola, Wendy, Xiang-Zhan and anyone else I might have missed.

Finally I must thank my family and friends for their support and encouragement, particularly Mum, Dad, Pete and Tom.

This thesis is dedicated to the memory of my Grandma, Peggy Grubb.

"Electronic calculators can solve problems which the man who made them cannot solve; but no government subsidised commission of engineers and physicists could create a worm."

Joseph Wood Krutch

An investigation of genetic heterogeneity in a biological sentinel species (*Lumbricus rubellus*)

Contents

Chapter 1. General introduction

As an introduction to the thesis it is perhaps worth briefly describing how the focus of my studies changed over the course of this PhD. The initial focus of this PhD was to investigate the possible existence of genetic adaptation within populations of the earthworm species *Lumbricus rubellus*. This particular species is widely applied as a sentinel species to monitor the ecological effects of contaminants within studies of soil ecotoxicology. As a sentinel species it is important that the species displays genetic homogeneity between its populations (Hilty and Merenlander 2000). This means that experimental findings based upon one population of the species can be considered representative of all other populations.

Concerns about the genetic adaptation of some *L. rubellus* populations to metal contamination have arisen following the demonstration of inherited resistance within populations from contaminated sites (Langdon et al. 1999, Langdon et al. 2001). The initial aim of this study was therefore to investigate the possible genetic adaptation of *L. rubellus* populations along a gradient of nickel contamination. This was attempted by conducting a population genetic analysis using microsatellite markers. Microsatellite analysis has indicated the possible existence of genetically adapted populations within studies of other species by revealing an increased degree of genetic differentiation between environmentally-stressed populations and unaffected populations (Plath et al. 2007, Tobler et al. 2008, Plath et al. 2010).

However, shortly after the start of this study it was discovered that the 'species' of *L. rubellus* may actually comprise two distinct cryptic genetic lineages (King et al. 2008). The existence of such lineages could potentially complicate any studies of metal resistance given that comparative lineages in other annelid species have been found to differ markedly in their tolerance of contaminants (Sturmbauer et al. 1999, Linke-Gamenick et al. 2000). Differences in tolerance can also affect the distribution of such lineages across contaminated and non-contaminated sites (Sturmbauer et al. 1999).

Following the discovery of these cryptic lineages within *L. rubellus*, the focus of the PhD was broadened into a wider consideration of genetic heterogeneity within the species that included studies of these two lineages. The first of these studies applied mitochondrial 'bar-coding' sequences to investigate the extent of cryptic variation within *L. rubellus* populations from sites located across Britain. This study also applied a morphological analysis in order to uncover lineage-specific traits that could potentially be applied to differentiate the two lineages. The existence of such morphological differences could enable the development of 'type specimens' for the two lineages (Schlick-Steiner et al. 2007).

The second study of cryptic variation applied the same microsatellite markers from the nickel study to investigate genetic differentiation between sympatric populations of the two lineages. This was conducted using individuals from two of the sites sampled during the nickel study. Mitochondrial genotyping had revealed both cryptic lineages to be present at these sites. By applying nuclear molecular markers it was therefore possible to determine whether the sympatric populations displayed any evidence supporting the existence of hybridisation between the lineages.

Finally the original study of the thesis, the population genetic analysis of *L. rubellus* along a nickel contamination gradient was conducted after genotyping the individuals to determine their lineage. It was decided to focus the study upon only a single lineage (lineage A) as this lineage was found to be prevalent at the most highly contaminated sites.

***Lumbricus rubellus*- general biology**

Within the field of soil ecotoxicology the earthworm species *Lumbricus rubellus* Hoffmeister (1843), has been widely applied as an important sentinel species (Spurgeon et al. 2003b, Fränzle 2006, Bundy et al. 2007). *L. rubellus* is a common lumbricid earthworm species that is easily recognisable on account of its dark-red pigmentation. It is often highly abundant within temperate soils and burrows horizontally through the upper litter-layer feeding on decaying organic matter (Sims and Gerard 1999). It is therefore classified as an epigeic species in accordance with the ecological categories of earthworm proposed by Bouché (1977). Given its epigeic ecological niche, *L. rubellus* largely favours the organically enriched soils found in habitats such as deciduous forests, flood plains, meadows and pastureland (Terhivuou 1988, Sims and Gerard 1999). However, *L. rubellus* can also often be found inhabiting soils in urban and industrialised areas (Pizl and Josens 1995).

L. rubellus displays a holarctic distribution and is widely distributed throughout Europe (Sims and Gerard 1999). Given that it is widely believed that the Pleistocene glaciations eradicated the earthworm fauna of Britain, it has been proposed that the present-day British population of *L. rubellus* may descend from populations that inhabited ice-free refugia in Southern Europe during the last ice age (King et al. 2008). The human-mediated dispersal of *L. rubellus* has also been proposed to have resulted in its present wide-ranging distribution. Many invasive populations of *L. rubellus* have been discovered in locations outside Europe including North America (Addison 2009), and Australasia (Schon et al. 2011).

L. rubellus is an out-crossing hermaphrodite species that displays sexual reproduction (Sims and Gerard 1999). As a hermaphrodite, the anterior portion of each adult individual features both

male reproductive organs such as the testes, testes sacs and spermathecae, and female reproductive organs such as the ovaries and ovipores (Edwards and Bohlen 1996). Being an oligochaete species, *L. rubellus* also displays a thickened collar of tissue termed the clitellum located after its 26th anterior segment (Sims and Gerard 1999). This organ is essential for the formation of earthworm cocoons. The breeding study of Elvira et al. (1996) found *L. rubellus* to display a maturity time of 74 days. *L. rubellus* also demonstrated a breeding rate of 0.43 cocoons a week. In comparison with the results from other epigeic species the breeding rate of *L. rubellus* could be considered relatively low (Elvira et al. 1996).

Transects have indicated that field distributions of *L. rubellus* and other epigeic earthworms can often be patchy with the existence of aggregations separated by gaps of several metres or more (Campana et al. 2002). Such aggregations may correspond to spatial differences in conditions such as organic matter content (Campana et al. 2002, Whalen 2004). As an example of this, aggregations of *L. rubellus* are often found underneath dung pats within pastureland (Sims and Gerard 1999). It has also been proposed that the formation of aggregations may aid reproduction within low dispersal species (Whalen 2004).

Estimates for the dispersal capacity of *L. rubellus* have largely been derived by observing population expansion within formerly earthworm-free soils. The study of Marinissen and van den Bosch (1992) which focused upon Dutch polder soils, revealed a dispersal rate of up to 12.8m year^{-1}. This value was far higher than the dispersal rates stated in the review of Eijsackers (2010) which were largely derived from populations colonising contaminated soils. Some of these populations displayed a dispersal rate of approximately 2m year^{-1}. It has also been proposed that *L. rubellus* may also be dispersed over larger distances by human activities. Such activities include the movement of soil through agriculture (Marinissen and van den Bosch 1992) and the disposal of earthworm bait by fishermen (Terhivuou 1988).

***Lumbricus rubellus* as a sentinel species**

Many of the previously mentioned characteristics of *L. rubellus* make it an ideal sentinel species for studies of soil contamination. These include its abundance within the soil environment, recognisable appearance and widespread distribution which all make it an easy species to collect during field studies. The low motility of *L. rubellus* also makes it an ideal sentinel species for conducting high resolution studies of soil contamination as individuals generally only accumulate contaminants from an area of approximately ten cubic metres (Fränzle 2006). Finally *L. rubellus* is also ecologically important given that like other earthworm species it is an "environmental engineer" of the soil environment (Spurgeon et al. 2003b) and an important prey-species for many animals (Fränzle 2006).

However, several studies have raised doubts about the suitability of *L. rubellus* as a sentinel species of soil contamination by indicating the possible existence of significant genetic heterogeneity within the species. Firstly breeding studies have indicated that some populations of *L. rubellus* from contaminated sites may display genetic adaptation towards heavy metal contaminants (Langdon et al. 1999, Langdon et al. 2001). The possible existence of metal-resistant populations of *L. rubellus* at contaminated sites has also been indicated by population genetic studies that have indicated genetic differentiation between exposed and non-exposed populations (Andre et al. 2010, Simonsen and Klok 2010). Secondly genetic heterogeneity has also been uncovered within *L. rubellus* populations in the form of cryptic genetic variation. The application of mitochondrial sequence analysis in the studies of King et al. (2008) and Andre et al. (2010) has uncovered evidence of two distinct cryptic lineages. The study of Andre et al. (2010) also uncovered evidence of genetic divergence within the nuclear genome.

The history of ecotoxicology

Over the past two centuries the development of heavy industry around the world has led to the release of a wide-range of different anthropogenic contaminants into the natural environment. This has occurred as a consequence of direct release into the global commons through activities such as the disposal of industrial and domestic waste, the direct release of agrochemicals, the indirect release of substances as by-products of mining and also the inadvertent release of large quantities of contaminants during accidents.

Some incidences of animal deaths following exposure to anthropogenic contaminants were reported during the industrial revolution of the nineteenth century (Hoffman et al. 1990). However, given that the science of ecology was still then in its infancy with important concepts such as food-chains and niches still to be discovered (Clapp 1994), the possible ecological effects of contamination received little attention.

During the latter half of the twentieth century following the impact of such influential publications as 'Silent Spring' (Carson 1963) and several well-publicised pollution incidents such as Lekkerkerk in the Netherlands and Love Canal in the USA (Pollard et al. 2001, Binns 2004), many countries began to introduce legislation requiring companies and organisations to assess the effects of their activities upon the environment (e.g. Department of the Environment 1990). Much of this legislation was based upon the findings of traditional 'dose/response' toxicological tests that were used to determine 'dangerous' levels of contaminants (Chapman 1995). However the assessment of ecological risks using traditional laboratory-based toxicological testing was increasingly seen as an over-simplification, with test species often selected primarily on the basis of cost or practicality rather than their ecological importance and with factors such as the effects

of different combinations of chemicals or different environmental conditions being largely ignored (Chapman 2003). Such concerns led to the development of a new form of toxicology termed 'ecotoxicology' (Butler 1978, Moriarty 1983, Walker et al. 1996). Ecotoxicologists sought to overcome the limitations of existing toxicological testing in assessing ecological risk by incorporating ideas and analytical methods from ecology. This included emerging ideas of community theory, life-history theory and population genetics (Van Straalen 2003). The development of ecotoxicology also led to a greater consideration of the 'functional end-points' of contaminants upon ecosystem performance (Van Straalen 2003).

One major problem in developing the field of ecotoxicology has however been in attempting to successfully determine the effects of contaminants upon ecosystem functioning. This is complicated by several factors both relating to knowledge of the contaminants and the ecosystem under investigation. Firstly the bioavailability of contaminants to a given ecosystem can change in accordance with several different factors such as the presence of other contaminants (Chapman et al. 2003), abiotic environmental factors such as pH (Peijnenburg et al. 1999, Lanno et al. 2004) and differences in the ability of organisms to bioaccumulate contaminants from the surrounding environment (Beeby 1991). Secondly assessments of ecosystem functioning can be complicated given the multitude of different ecological interactions that they may feature, many of which are often poorly understood or unknown to researchers (Anderson et al. 1994). Given these complexities it is clear that traditional toxicological testing methods with their reliance upon test-species unrepresentative of those in natural ecosystems may be poor monitors of ecological effects.

Sentinel species

One way by which ecotoxicologists have sought to monitor the effects of contaminants upon ecosystems has been by focusing upon one or more 'sentinel' species rather than undertaking a whole-ecosystem survey of contamination effects (LeBlanc and Bain 1997, Mhatre and Pankhurst 1997, Hilty and Merenlander 2000, Tabor and Aguirre 2004, Fränzle 2006). Although this approach has been criticized as an oversimplification (Power and McCarty 1997), the application of sentinel or indicator species to study the effects of habitat change has found favour as a cheaper and more efficient means of assessing ecosystem health than larger scale monitoring studies (Pearson 1995, Caro and Doherty 1998). Sentinel species can be applied both for precautionary laboratory-based simulation studies that aim to determine the possible effects of contaminant exposure upon the functioning of exposed ecosystems and also in field studies of actual contamination events (Francioni et al. 2007, Quirós et al. 2007).

In order for a sentinel species to yield results that are meaningful and relevant to the ecosystem and contamination event under investigation there are several important criteria that need to be fulfilled in the selection of a suitable species. First and foremost the species under investigation must be ecologically important within the ecosystem to which it belongs such that any detrimental effects of the contaminants upon the species either in terms of its functioning or demography could have repercussions at different trophic levels (Chapman 2002). Ecologically important species which may be assigned as sentinels include species that play a role in the decomposition of organic matter (Zaitsev and van Straalen 2001), prey species of various different trophic levels (Hamers et al. 2006) and 'ecological engineer' species which through their activities change the physical characteristics of their environment (Jouquet et al. 2005).

It is also vitally important that the sentinel species displays some degree of sensitivity towards the contaminant under investigation. However, given that a species that displays high mortality at low levels of contamination would be of little use for assessing the long-term chronic effects of contaminants, it is important that the species displays some degree of tolerance towards the contaminant.

If a species displays well-characterized stress responses towards contaminants, diagnostic 'biomarkers' can be developed for ecotoxicological studies (LeBlanc and Bain 1997, Eason and O'Halloran 2002, Vasseur and Cossu-Leguille 2003). Biomarkers can be broadly defined as 'any biological response to an environmental chemical at the individual level or below demonstrating a departure from normal status' (Eason and O'Halloran 2002). The ultimate aim of biomarker studies is to detect any negative effects of a contaminant upon individuals before damage is able to occur at the population-level (Forbes et al. 2006). The majority of biomarkers result from the biological change or damage of an organism following exposure to contaminants and are detectable by behavioural, biochemical, cellular, immunological, genetic or morphological analysis (Depledge et al. 1995; LeBlanc and Bain 1997).

An ideal biomarker would display several characteristics, namely a high specificity and sensitivity to the toxicant under investigation and the ability to rapidly indicate stress (Van Straalen and Roelofs 2008). Many biomarkers that fulfill such criteria have been uncovered following the development of ecotoxicogenomic analysis techniques, which include such 'omic' technologies as transcriptomics, proteomics and metabolomics (Van Straalen and Roelofs 2008, Poynton and Vulpe 2009, Fedorenkova et al. 2010). By identifying the genes and proteins that are expressed following exposure to a contaminant, ecotoxicogenomic studies are able to resolve the pathways underlying the stress-responses of an organism (Spurgeon et al. 2008, Van Straalen and Roelofs 2008, Poynton and Vulpe 2009). This can enable the development of biomarkers indicating cellular

damage or the activation of detoxification mechanisms (Bundy et al. 2008, Van Straalen and Roelofs 2008, Brulle et al. 2010).

Sentinel species may also be selected on the basis of the ease with which they can be collected from within an environment. Some of the criteria proposed for selection of sentinels are therefore that the species is abundant within the study area and easily identifiable (Caro and O'Doherty 1998). Given that even closely related species can differ markedly in their sensitivity to contaminants (e.g. Bach et al. 2005), the correct identification of a sentinel should also be viewed with importance.

Other factors that may be important in the selection of a sentinel include the home range and ecological niche of a species (Hilty and Merenlender 2000). Regarding home range, many field studies have sought to apply sentinel species of low mobility or in some cases sessile life histories. This is advantageous both in ensuring that the study species in unable to avoid exposure to the contaminant under investigation (Hilty and Merenlender 2000) and in enabling high resolution studies of the effects of contaminants within a given environment (Fränzle 2006). Studies applying species of low mobility may however be restricted by the natural field distribution of these organisms (Ji et al. 2006). In a similar manner to home range, the ecological niche of a species might also be important in determining the accuracy with which it indicates environmental stress. Species inhabiting a specialized ecological niche may be expected to face greater exposure to contaminants through their diet than species of a more generalist niche which may be able to adapt their diet to avoid the contaminants (Hilty and Merenlander 2000).

Earthworms as sentinel species

Within the field of soil ecotoxicology earthworm species have been found to represent important sentinel species (Spurgeon et al. 2004a, Spurgeon et al. 2004b Bundy et al. 2007, Owen et al. 2008, LaCourse et al. 2009). Many earthworm species display characteristics that make them ideal sentinels of soil ecosystems. Firstly the relatively large size of earthworm species in comparison with other soil invertebrates such as oribatid mites makes earthworm species convenient for field identification and collection without the requirement of magnification (Mhatre and Pankhurst 1997, Paoletti 1999). Earthworms may also be highly abundant within their respective environment with some areas featuring earthworm densities of several hundred individuals per square metre of ground (Edwards and Bohlen 1996, Paoletti 1999). Many earthworm species have also been well-studied in terms of their ecology (Bouche 1992, Sims and Gerard 1999), and can be broadly defined as epigeic (leaf-litter dwelling), endogeic (laterally burrow through mineral soil) and anecic (vertically burrow through mineral soil).

The functional importance of earthworm species within their respective ecosystems has also been well documented by many studies. Many earthworm species have been found to be significant prey species for generalist predators (Symondson et al. 2000) and as such form an integral part of their respective food webs. Earthworms also play a prominent role within their ecosystem as 'engineers' of the soil environment (Lavelle et al. 1997, Jouquet et al. 2006, Turbé et al. 2010). Given that through their activities earthworms convert organic matter into inorganic compounds and move large quantities of soil to the surface, earthworms are directly involved in the cycling of soil nutrients. They may also play a more indirect role in nutrient cycling by enabling the dispersal of important decomposing fungi and bacteria (Brown and Doube 2004). The burrows of anceic and endogeic earthworm species can also be important in increasing the infiltration of water into soil with obvious implications for plant growth (Edwards and Bohlen 1996, Cannavacciuolo et al. 1998). Earthworm activity may also stimulate plant growth by increasing the expression of stress-responsive genes, increasing the tolerance of plants towards environmental stressors (Blouin et al. 2005).

A second major characteristic of earthworms that qualifies them as ideal sentinel species is their sensitivity towards a wide range of different soil contaminants (Owen et al. 2008, Brulle et al. 2010). This sensitivity is indicated by ecological surveys which have uncovered significant reductions in earthworm abundance and species diversity at contaminated sites (Nahmani et al. 2003, Lukkari et al. 2004). The physiology and behavioral aspects of earthworm life-cycles pre-disposes them towards contaminant exposure. Earthworms may become exposed to contaminants through pore water uptake or the ingestion of contaminated soil (Osté et al. 2001, Vijver et al. 2003, Morgan et al. 2004). There has been some debate as to which of these uptake routes may play the most significant role in the absorption of contaminants. The study of Vijver et al. (2003) applied oral sealing to contaminant-exposed earthworms revealing little change in contaminant uptake. This suggested that the uptake of contaminants through pore water could be more significant than that occurring from ingested soil. However the study of Osté et al. (2001) modified the availability of contaminants through pore water, revealing little difference in their accumulation by earthworms. This contrastingly suggested that gut uptake could be the major uptake route. It is possible that the availability of contaminants by different uptake routes may however depend on the predominant chemical form of the contaminants within the soil (Nanonni et al. 2011).

Comparison of earthworm body burdens of contaminants together with environmental concentrations often reveals a linear rate of accumulation (Ireland 1983, Morgan and Morgan 1988). This occurs chiefly for the body concentration of unregulated non-essential elements (Nanonni et al. 2011). Some earthworm species may also display the ability to bioaccumulate

certain contaminants within their body tissues to concentrations far exceeding those of the surrounding environment. This is particularly true of the bioaccumulation of heavy metals by some earthworm species that are able to sequester large quantities of metals into chloragogoneous tissue surrounding their gut wall (Morgan and Morgan 1998, Carpené et al. 2006). It has been proposed that the bioaccumulation of heavy metals in these particular species may result from a physiological requirement for certain essential trace metals, with the heavy metals in question being chemically similar to a required metal (Beeby 1991).

Laboratory-based exposure trials have revealed that many earthworm species also display well-characterized stress responses upon exposure to even small quantities of contaminants. This has enabled the development of a suite of different biomarkers for ecotoxicological studies (for a review of these see: Kammenga et al. 2000; Scott-Fordsmand and Weeks 2000; Spurgeon et al. 2003). These biomarkers encompass a range of distinct stress responses that can be grouped more broadly into several distinct categories, namely cellular based assays (e.g. Booth and O'Halloran 2001; Sorour and Larink 2001; Jensen et al. 2007), detoxifying enzymes and proteins (e.g. Booth and O'Halloran 2001; Łaszczyca et al. 2004; Morgan et al. 2004), exposure related molecular changes not connected to detoxification (e.g. Nadeau et al. 2001) and metabolomic/proteomic profiles (e.g. Bundy et al. 2007, Bundy et al. 2008). Research has also focused upon the development of genetic biomarkers that indicate the changes in gene expression underlying physiological stress-responses (Galay-Burgos et al. 2003, Spurgeon et al. 2004, Asensio et al. 2007). This has led to the development of the internet-based database of earthworm expressed sequence tags (ESTs), LumbriBASE (www.earthworms.org, Stürzenbaum et al. 2003).

Some biomarker studies of earthworm species reflect stress-responses that could have direct demographic effects upon earthworm populations. These include a reduction in cocoon production (Booth and O'Halloran 2001) and cellular damage to spermatozoa (Reinecke and Reinecke 1997, Sorour and Larink 2001). Other biomarkers including cellular integrity and metallothionein gene expression have been correlated with changes in levels of organic matter removal by earthworm communities (Spurgeon et al. 2005). Biomarkers may also indicate the existence of environmental stress within earthworm populations before demographic effects are able to occur (Spurgeon et al. 2005).

Genetic homogeneity and sentinel species

An important characteristic for an effective sentinel species is that it displays genetic homogeneity between populations (Hilty and Merenlender 2000). This allows inferences of pollution tolerance that have been determined using only a small number of populations to be applied to other populations of that species. Unfortunately many studies have revealed that an assumption of

genetic homogeneity may be false for many sentinel species. A lack of homogeneity may occur for two different reasons, namely the adaptation of some exposed populations to contaminants (Posthuma and Van Straalen 1993) and the occurrence of cryptic genetic lineages within a 'species' (Bickford et al. 2006). Such genetic heterogeneity is of concern to ecotoxicologists as the existence of unrecognised genetic variation within sentinel species could compromise their ability to produce reliable indications of environmental stress. The susceptibility to contaminants of some natural populations of a genetically heterogeneous species could differ from that established by laboratory testing, undermining any attempts to establish 'safe' levels of contaminants (Evenden and Depledge 1997).

However some researchers have proposed that the incorporation of genetically variable species into environmental risk assessment trials may increase the accuracy of results (Forbes et al. 1998, Coe et al. 2009). Genetically diverse sentinels could produce results that are more applicable to wild populations (Coe et al. 2009).

Studies indicate that the populations of many earthworm species may exhibit genetic heterogeneity that could complicate their appplication as ecotoxicological sentinels. One such source of heterogeneity is the genetic adaptation of earthworm populations to soil contamination. Many studies have demonstrated that earthworms collected from such populations demonstrate a greater degree of tolerance towards contaminants than do non-exposed individuals (Rozen 2006, Arnold et al. 2008, Langdon et al. 2009, Fisker et al. 2011). Laboratory breeding studies have indicated the inheritance of tolerance within some of these populations (Rozen 2006, Langdon et al. 2009, Fisker et al. 2011). This suggests that some earthworm populations may become resistant through a process of genetic adaptation.

The application of molecular markers to earthworm taxonomy has also uncovered a second major source of genetic heterogeneity within earthworm species namely cryptic genetic variation. Genetic analysis has uncovered evidence of genetically distinct lineages within many earthworm species (King et al. 2008, Pérez-Losada et al. 2009, Andre et al. 2010, James et al. 2010, Novo et al. 2010b). A further analysis of some of these 'species' has indicated the presence of reproductive isolation between lineages, which may suggest that they represent cryptic species (Lowe and Butt 2008, Dupont et al. 2011). Such distinct lineages could confound ecotoxicological studies given that cryptic lineages within some annelid species have been found to differ widely in their tolerance of contaminants (Sturmbauer et al. 1999, Linke-Gamenick et al. 2000).

Genetic adaptation

Studies of organisms inhabiting contaminated sites have repeatedly demonstrated the existence of tolerant populations. The individuals of these populations display comparatively higher levels of

growth and fitness, and lower mortality rates upon exposure to contaminants than do the individuals from non-exposed populations (Posthuma and Van Straalen 1993, Langdon et al. 2003, Eränen et al. 2009). Individuals of exposed populations have been demonstrated to display evidence of tolerance resulting from a reduced uptake of contaminants (Xie and Klerks 2004), an increased capacity for the detoxification of harmful compounds (Babczyńska et al. 2006, Gonzalez-Rey 2007, Goto and Wallace 2010) or increased excretion ability (Posthuma et al. 1992, Postma and Groenendijk 1999). Evidence of an increased tolerance towards contaminants has been uncovered across a diverse range of different organisms including plants (Jiménez-Ambriz et al. 2007, Eränen et al. 2009), and many different animal phyla (reviewed in: Posthuma and Van Straalen 1993) including annelids (Klerks 2002, Langdon 2003), arthropods (Diogo et al. 2007, Khan et al. 2011) and vertebrates (Klerks and Moreau 2001, Xie and Klerks 2004). Breeding trials have indicated that the tolerance observed within many of these species may have a heritable basis. The fact that one of the previously mentioned criteria for the selection of sentinel species is a limited home-range, may actually predispose some of these sentinels towards the development of genetic adaptation. The isolation of populations undergoing selection has been proposed as an important factor in enabling the development of adaptation, with high levels of gene flow into stressed populations possibly constraining divergence by introducing a high proportion of non-adapted genotypes (Hendry et al. 2002). Artificial laboratory based experiments of selection featuring individuals collected from 'clean' locations have also indicated that resistance may evolve rapidly within some species following only a few generations of exposure (Klerks and Levinton 1989).

It has also been demonstrated that different populations of a single species may differ in their ability to develop genetic resistance towards contaminants (Al-Hiyaly et al. 1988, Al-Hiyaly 1993). Such studies have indicated that an important prerequisite for the development of genetic adaptation is a high level of local genetic diversity, as such populations would be expected to include a higher proportion of resistant genotypes. The studies of Al-Hiyaly (1988) and Al-Hiyaly (1993) were able to demonstrate this by correlating the presence of zinc-resistant plant populations around pylons with the existence of resistant genotypes in nearby uncontaminated meadows. Alternatively it has also been proposed that populations may differ in their ability to adapt to stress if there is a high metabolic cost of maintaining genetic tolerance or negative pleitropic effects (Posthuma and Van Straalen 1993, Van Straalen and Timmermans 2002). In such situations exposed populations may fail to develop genetically based tolerance, instead displaying resistance through physiological acclimation (i.e. phenotypic plasticity) (Meyers and Bull 2002, Schlichting and Smith 2002, Roelofs et al. 2010, Whitman and Agrawal 2011). The existence of such variation in the capacity of populations to adapt to contamination could create mistaken impressions of the overall susceptibility of a species towards a given contaminant. If the sensitivity

of a species towards a given contaminant is established by sampling non-resistant populations, this may lead to an underestimation of the ability of other populations to tolerate the contaminant (Evenden and Depledge 1997). More worryingly though, the unsuspected sampling of genetically adapted populations could create a false indication of the adaptability of a species to tolerate contaminants.

It has been proposed that adapted populations may be identifiable on the basis of several demographic characteristics that may differentiate them from non-adapted populations (Posthuma and Van Straalen 1993, Belfiore and Anderson 2001, Van Straalen and Timmermans 2002). By analysis of the population genetics of exposed and non-exposed populations it may therefore be possible to infer the presence of genetically adapted populations at contaminated sites. Some advantages of this approach to the inference of genetic adaptation are that it can enable the identification of past historical demographic events (e.g. genetic bottlenecks) and also that it offers a greater degree of resolution in its inferences by focusing upon an entire population rather than the reactions of several individuals (Belfiore and Anderson 2001).

One way by which population genetic analysis may imply the presence of an adapted population is by indicating a low level of gene flow between populations within stressful environments and adjacent non-exposed populations. Levels of gene flow into a contaminated site may be reduced if strong selection increases the rate of mortality or reduces the fitness of the non-adapted migrant individuals (Slatkin 1987). Laboratory based studies comparing the survival and fitness of individuals from adapted and non-adapted populations upon exposure to strong physiological stress often indicate a comparatively higher mortality rate or lower fitness of non-adapted individuals (e.g. Langdon et al. 2003, Lopes et al. 2006). This can result in a low rate of successful immigration into adapted populations. It has also been suggested that in certain situations the successful migration of adapted individuals out of their environment may be reduced by 'costs of tolerance' (McNeilly 1968, Posthuma and Van Straalen 1993). Studies have demonstrated that adapted individuals may display a lower level of fitness than non-adapted individuals in low-stress environments due to factors such as the energetic cost of maintaining adaptation mechanisms (McNeilly 1968, Xie and Klerks 2004). Reduced levels of migration from non-adapted populations are believed to aid in the fixation of adapted genotypes within stressed populations by preventing 'gene-swamping' which can occur in situations of high gene flow and reduce the frequency of adaptive genes within the stressed populations (Lenormand 2002).

However, in some situations selection may have less of an effect upon gene flow particularly if the strength of selection is moderate or weak (Morgan et al. 2007). This has been reflected by studies demonstrating the existence of mixtures of adapted and non-adapted genotypes within stressed and nearby reference populations (Lopes et al. 2004). It has been proposed that in some

situations gene flow may actually assist in adaptation by introducing new adaptive alleles into stressed populations and by reducing the detrimental effects of inbreeding (Lenormand 2002, Alleaume-Benharira et al. 2006).

If strong selection upon a population has resulted in adaptation though, a reduction of gene flow may be identifiable between populations due to the development of neutral genetic changes within adapted populations. The genetic drift of allele frequencies within an adapted population can cause it to display genetic differentiation from nearby populations. By observing allele frequencies and comparing them between populations it is therefore possible to identify the effects of genetic drift within populations allowing an inference of reduced levels of gene flow between them. Studies may also apply an analysis of isolation of distance or observe the effects of geographical barriers upon such patterns of differentiation in order to distinguish genetic structuring through adaptation from that arising due to geographical isolation. One of the most widely applied statistical methods for identifying population differentiation is hierarchical F-statistics (Weir and Cockerham 1984).

It has also been proposed that populations experiencing strong selection may undergo 'genetic erosion', resulting in a significant loss of overall genetic diversity (Hebert and Luiker 1996, Bickham et al. 2000, Van Straalen and Timmermans 2002, Nowak et al. 2009). Such a loss of genetic diversity may occur within an exposed population without adaptation due to a significant reduction in population size, as has been recorded during the field studies of some exposed populations (e.g. Wilson et al. 2005). However the adaptation of a population could also cause a loss of genetic diversity for two different reasons. Firstly the selection of adapted individuals over non-adapted individuals either due to the increased mortality of non-adapted individuals or the greater reproductive success of adapted individuals could lead to an overall loss of genetic diversity (Bickham et al. 2000, Van Straalen and Timmermans 2002). This can occur due to a genetic bottleneck if a small number of adapted individuals are able to display a disproportionally high degree of reproductive success resulting in a population expansion (Bickham et al. 2000, Van Straalen and Timmermans 2002). Another alternative process by which adaptation can lead to a reduction in levels of genetic diversity is by the process of genetic hitch-hiking (Hebert and Luiker 1996, Schlötterer and Wiehe 1999, Belfiore and Anderson 2001). Genetic hitchhiking can occur when genes enabling adaptation to a particular environmental stress become fixed within a population. This can lead to a reduction in the levels of genetic polymorphism within the areas flanking the selected genes. The size of the area affected and therefore the degree of genetic hitch-hiking that may occur is dependent on both the strength of selection upon the allele and the recombination rate within its chromosomal region (Schlötterer and Wiehe 1999). Genes with a

high selection coefficient within areas of low recombination would be expected to result in the highest degree of genetic hitch-hiking.

However, there are two alternative explanations to adaptation for the existence of populations within stressful environments (Fig. 1.1), namely phenotypic plasticity or source-sink patterns of migration. Nevertheless it is possible to differentiate occurrences of both of these processes from genetic adaptation due to the different impacts upon population genetics that both would present.

Firstly, phenotypic plasticity refers to situations in which individuals of a single genotype are able to adapt to a range of environmental stresses by several different processes. These processes include physiological plasticity where individuals change their physiology or behavior to increase their tolerance towards an environmental stress, and developmental plasticity where juvenile individuals exposed to an environmental stress undergo developmental changes that make them more tolerant as adults (Meyers and Bull 2002, Schlichting and Smith 2002, Roelofs et al. 2010, Whitman and Agrawal 2011). Such phenotypic adaptations can occur due to changes in transcription following exposure to stress (Schlichting and Smith 2002). The emerging field of epigenetics has also indicated several mechanisms by which phenotypic changes could occur in response to stress. These include DNA methylation, histone modification and gene silencing by non-coding RNA (Vandegehuchte and Janssen 2011). It has been proposed that plasticity may occur within organisms inhabiting a naturally heterogenous environment (Crispo 2008) or as a precursor to the development of genetic adaptation (Schlichting and Smith 2002). If the individuals within a stressful environment are able to tolerate stress by exhibiting a phenotypically plastic genome, this would not be expected to impede the migration of individuals between stressed and non-affected populations as the effects of selection would be dampened (Fig. 1.1) (Crispo 2008). Consequently many studies of phenotypically plastic species have uncovered evidence of low levels of genetic differentiation between otherwise phenotypically distinct populations (Jordan et al. 2005, Whiteley et al. 2009). It may therefore be possible to differentiate between incidences of phenotypic adaptation (i.e. by physiological or developmental acclimation) and genetic adaptation by observing whether the populations of a stressed environment appear to be genetically homogenous or differentiated from one another, with several studies having uncovered possible incidences of phenotypic plasticity in this manner. Such studies include that of Jordan et al. (2005) that demonstrated the existence of gene flow between populations of a lizard from two different island habitats. This was in spite of the existence of habitat-specific differences in predator avoidance behavior between populations. Another study that was able to infer the existence of phenotypic plasticity was that of Whitelely et al. (2009)

who found no evidence of genetic differentiation between the separate populations of a fish species displaying substrate-specific body colouration.

The other possible explanation for the existence of populations within stressful environments is that they represent the 'sink' population within a 'source-sink' system. Such a population would be comprised of physiologically 'stressed' individuals that displayed a low degree of fitness and a high mortality rate. This 'sink' population would be maintained by a migration of individuals from other 'source' populations. Some studies have demonstrated evidence of source-sink systems by inferring high levels of gene flow between 'sink' populations and their possible 'source' populations. The separate source populations may also be significantly genetically differentiated from one another reflecting the asymmetrical nature of gene flow out of these populations. Sink populations may also be identifiable through a high degree of genetic diversity as a result of them being sustained by migrants from separate populations. Studies that have inferred the existence of possible source-sink dynamics include that of Waits et al. (2008) that indicated that fish populations within a degraded urban waterway may be sustained by migration from upstream populations. The study of Dias et al. (1996) was also able to infer the existence of a source-sink dynamic between populations of birds inhabiting deciduous and evergreen forest habitats. In this later study, evidence of a far lower degree of genetic differentiation was found between suspected source and sink populations than between source populations and between sink populations.

Cryptic variation

A second way by which the genetic homogeneity of sentinel species may be compromised is through the existence of unrecognised cryptic genetic diversity. Molecular analysis has already uncovered significant levels of hitherto unrecognised cryptic diversity and genetic lineages within many invertebrate species including annelids (Sturmbauer et al. 1999, Erséus and Gustafsson 2009), molluscs (Baker et al., 2003), crustaceans (Adamowicz et al. 2007, Wares et al. 2007) and insect species (Chenon et al. 2000, Schlick-Steiner et al. 2006). Some of these species have been applied as sentinels within ecotoxicological studies; e.g. *Tubifex tubifex* within freshwater ecosystems (Sturmbauer et al. 1999), *Folsomia candida* within soil ecosystems (Chenon et al. 2000; Diogo et al. 2007) and *Idotea balthica* within intertidal ecosystems (Wares et al. 2007).

Several studies of sentinel species that are known to contain differentiated genetic lineages have uncovered evidence for variation in the levels of resistance to contaminants between these lineages. Distinct genetic variation was found between laboratory populations of the springtail, *Folsomia candida*, a species that has been widely applied in ecotoxicological studies of soil ecosystems (Diogo et al. 2007). Experimental analysis uncovered a distinct difference in the ability

of individuals from two genetic lineages of the species to tolerate pesticide contamination. In a behavioural trial individuals of one lineage were found to show a strong avoidance of soil contaminated with low levels of pesticide. However, individuals of the second lineage were found to show no avoidance behaviour when presented with the same soil. Differences in contaminant tolerance were also found between cryptic lineages of the freshwater oligochaete *Tubifex tubifex*, which is widely used as a sentinel species in freshwater ecosystems (Sturmbauer et al., 1999). Molecular analysis of *Tubifex tubifex* uncovered evidence of five distinct genetic lineages within European river populations. Toxicity experiments revealed variation in the levels of cadmium-resistance displayed by different genetic lineages. Geographical variation in cadmium tolerance between *T. tubifex* populations was found to be largely linked to different frequencies of genotypes within the populations. Finally, the marine polychaete species *Capitella capitata* has been demonstrated to actually represent a cryptic species complex, with genetic studies and analysis of reproductive traits confirming the existence of several reproductively isolated lineages (Grassle and Grassle 1976). Exposure trials featuring polyaromatic hydrocarbons revealed differences in sensitivity to toxicants to exist between these lineages (Linke-Gamenick et al. 2000). Further analysis indicated the more tolerant lineages to display a greater capacity for biotransformation of contaminants than the more sensitive lineages (Bach et al. 2005). The existence of unrecognised genetic heterogeneity within a 'species' could therefore confound its application in ecotoxicology.

Molecular markers

This study will apply two different types of molecular marker to ascertain the degree of genetic heterogeneity within *L. rubellus*, namely mitochondrial DNA and microsatellite regions. The first of these molecular markers, mtDNA is a particularly useful molecular marker for the definition of recently diverged taxa as it displays an evolutionary rate that is far higher than most nuclear genes (Avise 2000). This also results in high levels of intraspecific genetic polymorphism that enables the discovery of intraspecific genetic lineages. The tracing of genetic lineages is also helped by the asexual, maternal mode of inheritance[1] displayed by mtDNA that assists in the definition of genetic lineages by removing the confounding effects of recombination.

Genetic variation between mitochondrial genes largely occurs in the form of nucleotide substitutions or minor variations in length, rather than by changes in gene order (Avise 2000). The occurrence of insertion or deletion (indel) mutations can however create problems during taxonomic analysis by complicating sequence alignments between different haplotypes. This has led to the protein-coding genes of mitochondrial DNA being favoured for broad taxonomic studies

[1] As *Lumbricus rubellus* is an out-crossing simultaneous hermaphrodite, the terms maternal and paternal here refer to the lineages represented by gametes rather than individual earthworms.

over genes that display a high incidence of indel mutations such as the rDNA genes, 12S and 16S (Hebert et al. 2003). Selection against frame-shift mutations results in a low incidence of indel mutations within protein-coding genes.

The development of robust primers has allowed mtDNA analysis to be applied over a wide variety of different animal phyla. The mitochondrial cytochrome oxidase I (COI) primers (Folmer et al. 1994) are a general invertebrate tool that has been widely applied in many phylogenetic studies (Bely and Wray 2004, Williams et al. 2006, Adamowicz et al. 2007). The primers were designed using mtDNA sequences drawn from species ranging across several invertebrate phyla. The COI primer pair amplify a 710b.p. fragment of the cytochrome c oxidase subunit I gene. This gene encodes a part of the enzyme that is important in cellular respiration. Consequently the COI sequence is highly conserved across phyla, accounting for the wide applicability of the COI primer pair. The conserved nature of the sequence also means that its rate of evolution is lower than that of non-coding mtDNA. This lower rate of evolution has made the COI gene a favoured molecular marker for the construction of species phylogenies (Damgaard et al. 2008; Guan et al. 2008; Mengual et al. 2008).

The wide-ranging applicability of the COI primer pair led to proposals that the COI sequence could be used as a genetic 'barcode' for the diagnosis of species (Hebert et al. 2003; Hebert et al. 2004). This was partly based upon the finding that levels of COI sequence variation within taxonomic species are generally far lower that those found between taxonomic species (Hebert et al. 2004). Species phylogenies based upon mtDNA sequence variation had already been found to be widely convergent with traditional taxonomies (Avise and Walker 1999). This agreement between the findings of mtDNA analysis and taxonomy work has led to COI analysis being widely applied as a means of uncovering potential cryptic species sometimes as a precursor to more detailed studies of taxonomic or ecological differences (Schlick-Steiner 2006; Adamowicz et al. 2007; Wares et al. 2007). The mtDNA COI sequence will therefore be applied within this study to observe cryptic variation within *L. rubellus*.

The second molecular markers that will be applied within this study, microsatellite loci are found throughout the nuclear genome and have been proposed by some researchers to represent evolutionarily neutral markers (Schlötterer and Wiehe 1999, Chistiakov et al. 2005). Such characteristics mean that they are not constrained by some of the limitations associated with mtDNA markers such as maternal inheritance and genetic linkage (Ballard and Whitlock 2004). Microsatellites consist of repetitive segments of DNA and display a high degree of size variation caused by differences in the number of these repeats (Hancock 1999). Such size differences are believed to develop through the process of replication slippage (Eisen 1999, Calabrese and Sainudiin 2004). Replication slippage occurs when the template and copied strands of DNA

become misaligned during replication and can lead to the loss or gain of repeat units. The accumulation of microsatellite mutations through replication slippage can occur far more quickly than for other processes such as point mutation (Hancock 1999) and consequently reproductively isolated populations can diverge in allele frequencies over a relatively short time period. This has made microsatellite loci popular molecular markers for the resolution of fine-scale demographic structuring (e.g. Van der Wurff et al. 2003, Plath et al. 2007, Tobler et al. 2008, Plath et al. 2010). Given the sensitivity of microsatellite markers to demographic changes, they will be applied within this study to observe whether populations of *L. rubellus* display adaptation to soil contamination. Several microsatellite markers have already been developed for use with *L. rubellus* (Harper et al. 2006) and the congeneric species *L. terrestris* (Velavan et al. 2007).

Given that microsatellite allele frequencies can diverge between the isolated populations of a single species, some studies have sought to apply allele frequency data to infer the existence of reproductive isolation between the populations of separate species (Feulner et al. 2006, Wares et al. 2007) or to uncover evidence of interspecific hybridization (Luo et al. 2010, Väli et al. 2010). Many of the interspecific studies that have applied microsatellite frequency data to their analyses have employed the same statistical approaches as population genetic studies. This includes the determination of population differentiation through inbreeding-coefficients such as F-statistics to infer the degree of gene flow between populations. This approach has successfully demonstrated the existence of reproductive isolation between species of bats (Racey et al. 2007), fish (Feulner et al. 2006) and arthropods (Li et al. 2009). Interspecific studies have also applied exploratory data analysis such as Bayesian clustering methods to observe the degree of reproductive isolation that exists between sympatric populations of different species. This approach has been particularly useful for studies aiming to uncover evidence of hybridization uncovering evidence of introgression between species of plants (Zeng et al. 2010), molluscs (Luo et al. 2010) and birds (Väli et al. 2010).

However, it has been proposed that microsatellites may display some complexities in their process of mutation that could complicate such studies of interspecific genetic differentiation. Initially it was assumed that microsatellites develop through a stepwise mutational model with new microsatellite alleles arising through slippage mutations that either add or subtract single repeat units (Calabrese and Sainudiin 2004). However, it has since been proposed that microsatellites may sometimes lose or gain multiple repeat units within a single mutation, leading to suggestions that a 'two-phase' model may more accurately reflect the actual mutation of microsatellites (Calabrese and Sainudiin 2004). The adoption of a linear model of mutation may also be regarded as an over-simplification as patterns of mutation are believed to change with differences in allele length. Larger microsatellite repeats are believed to undergo slippage

mutations more frequently and consequently display a higher mutation rate than shorter microsatellites (Calabrese and Sainudiin 2004). It has also been proposed that large microsatellites may face size constraints preventing further expansion and consequently display a higher rate of deletion mutations (Nauta and Weissing 1996). The interspecific comparison of some conserved primate microsatellite loci revealed a far lower than expected degree of size variation to exist between the alleles of some human and chimpanzee microsatellites which suggested the existence of size constraint upon these loci (Garza et al. 1995).

The sequencing of microsatellite flanking and repeat regions has revealed many alleles of the same size to be representative of different evolutionary lineages that have converged in repeat number (Dettman and Taylor 2004, Domingo-Roura et al. 2005). An extreme example of this was uncovered by the study of Domingo-Roura et al. (2005) who uncovered 28 species-specific microsatellite alleles across 23 different species of carnivore after sequencing nine different size alleles from across these species. Such homoplasy is a consequence of the fact that microsatellites can mutate by both expansion and deletion of repeat units. As has already been stated, larger microsatellite alleles are believed to face constraints upon expansion and display an increased rate of deletion. Over large evolutionary time-scales therefore the microsatellite alleles within divergent species may be expected to converge in allele size, obscuring patterns of phylogenetic descent.

Nevertheless, despite the existence of these possibly confounding problems, several studies have managed to successfully apply microsatellite allele data to identify species and in some cases to establish a phylogeny of descent. At the most basic level, some studies have identified species on the basis of disjunctive microsatellite size-ranges (Molbo et al. 2003, Elmer et al. 2007). Many more studies have established the existence of reproductive isolation between populations of species by applying analysis methods common to population genetics including population inbreeding coefficients and clustering analysis. One such study that was able to successfully apply this methodology was that of Petren et al. (1999), who utilised microsatellites to resolve the phylogeny of the Galapagos finch island radiation. The study was able to resolve many species that had previously been defined on the basis of morphological or allozymic differences. However, the study was also able to identify the existence of reproductive isolation between several morphologically indistinct island populations of the 'species' *Certhidea olivacea*, indicating the existence of several hitherto overlooked cryptic species.

An alternative approach for interspecific studies of microsatellites has been uncovered by the sequencing of some cross-amplifying microsatellite loci. As has already been mentioned some cross-taxa studies of microsatellites have applied sequencing of the microsatellite flanking region (MFR) to determine the extent of size homoplasy at these loci (Dettman and Taylor 2004,

Domingo-Roura et al. 2005). These studies have uncovered evidence of species-specific differences in point mutations and insertion/deletion mutations within the flanking region of microsatellites. Estimates of the mutation rate of MFRs have been determined by comparing sequences from closely related species (Brohede and Ellegren 1999). Such studies have uncovered mutation rates that are close to the values stated for known introns. This indicates that MFRs can be considered evolutionarily neutral, and that the high degree of sequence conservation in the primer binding region is not a result of selection. Several studies have successfully applied sequence data derived from the MFRs of highly conserved microsatellite loci to infer phylogenies across a variety of different evolutionary scales. These studies range from the phylogenetic resolution of closely-related congeneric species groupings (Dettman et al. 2003, Feulner et al. 2006) to the inference of phylogenies for more evolutionarily distant taxa such as families (Domingo-Roura et al. 2005).

One of the first studies to directly analyse the interspecific sequence variation of microsatellite loci was that of Schlötterer et al. (1991), who observed sequence differences between whale species. In this study several microsatellite primer pairs originally developed for studies of the long-finned pilot whale *Globicephala melas*, were found to successfully cross- amplify eleven other whale species from across two sub-orders that are believed to have diverged 35-40 Mya. Interspecific comparisons of these microsatellite sequences revealed the presence of single nucleotide polymorphisms (SNPs), both within the repeat regions and to a lesser extent, within the flanking regions. A relatively constrained rate of sequence mutation within these regions was indicated by the observation that many SNPs appeared to have deep evolutionary origins, with some mutations being shared across several species and therefore appearing to have their origins in mutational events predating the divergence of these species. As would be expected for microsatellites, between species there was a great deal of evidence for the occurrence of insertion and deletion mutations within the repeat region. Such mutations were largely the result of the loss or gain of one or a few repeat units and were therefore most likely to have resulted from slippage mutations rather than to be the result of isolated mutational events. However, evidence of substitutional mutations was found within the repeat regions of all of the microsatellites that were investigated. As with the mutations within the MFR, some of the substitutions were found to be conserved across several closely related species suggesting a relatively deep evolutionary origin.

The studies of Zardoya et al. (1996) and Rico et al. (1996), both applied sequence analysis to study interspecific differences in microsatellites conserved over an evolutionary time period even greater than that of the cetacean microsatellite studied by Schlotterer. Both studies were able to apply sequencing analysis to uncover differences in conserved microsatellites between fish

species that diverged from one another >80 Mya. In the study of Rico et al. (1996) this included fish species from three separate classes. Both of the studies uncovered evidence supporting the occurrence of large insertion/deletion mutations at deeper taxonomic levels. In the study of Rico et al. (1996) such mutations were found predominantly between fish species of different classes. However, in their microsatellite analysis of perciform fish, Zardoya et al. (1996) uncovered insertion/deletion mutations in the microsatellite flanking regions that were found to be unique to particular families of Labroid fishes. Indel mutations were also discovered within the Cichlidae family that could be applied to differentiate Neotropical species from those of Africa and Asia. Interspecific differences in the microsatellite flanking region due to base substitutions were found across all of the taxonomic levels featured in both studies.

The studies of Blanquer-Maumont and Crouau-Roy (1995) and Domingo-Roura et al. (2005) both applied microsatellite sequence analysis for the resolution of complex phylogenies that had previously been investigated using conventional nuclear and mitochondrial coding sequences. The study of Domingo-Roura et al. (2005), applied sequence analysis of a highly conserved microsatellite region to study carnivore species from five different families. Many of the base substitutions and deletion mutations that were uncovered within the microsatellite flanking region appeared to persist across evolutionary time-scales, being shared by two or more species. The MFR mutations were applied to produce a phylogenetic tree which was found to be highly congruent with existing phylogenetic studies in its designation of the carnivore families.

A suitability of microsatellite flanking region sequences for the resolution of species has also been indicated by comparisons of levels of sequence variation between different congeneric species with the levels of variation between different populations of a single species. The study of Makova et al. (2000) uncovered a small degree of intraspecific sequence variation within the MFR of a mouse species. The different haplotypes of this MFR were found to correspond to two distinct groups of size alleles within the repeat region. A comparison of MFR alleles from different populations of the mice revealed no evidence for the population sorting of these alleles, with many shared between geographically distant populations. A comparison of MFR sequences across genera did however show them to form monophyletic clusters corresponding to existing species groupings. These phylogenetic inferences were found to be supported by an independent analysis of mtDNA sequence data. This indicates that the MFRs display a rate of sequence mutation that is comparable to the mitochondrial sequences that are often applied to resolve species in phylogeny studies.

Within this study the two lineages of *L. rubellus* will be analysed using microsatellite markers through both an analysis of allele frequencies and a sequence analysis of the microsatellites and their flanking regions.

Aims of the project

In this thesis the extent of genetic heterogeneity within *L. rubellus* will be analysed in three different studies that will apply molecular markers to uncover evidence of cryptic lineages and genetically adapted populations within the species.

The first of these studies will be a mitochondrial analysis using barcoding primers which will aim to further investigate the extent of lineage diversification within *L. rubellus*. Several populations will be analysed from sites located throughout Great Britain and individuals from three other *Lumbricus* species will also be investigated in the study. A morphological analysis will also be conducted to uncover any lineage-specific traits that could possibly be used to differentiate individuals of the two lineages of *L. rubellus*.

In the second study microsatellite analysis will be applied to analyse differences between the two lineages of *L. rubellus*. This analysis will include an investigation of allele frequencies that will apply population differentiation approaches to assess the degree of genetic isolation between the two lineages. Secondly, microsatellite regions will also be sequenced to observe possible sequence differences in the flanking and repeat regions. Past studies have indicated that such sequence differences are often found between divergent species. This study will aim to test the following hypotheses:

1. The two lineages of *L. rubellus* are reproductively isolated and more genetically differentiated than are allopatric populations of both lineages.

2. The lineages display evidence of genetic introgression with the existence of hybrid individuals

3. Microsatellite sequences display a greater degree of divergence between lineages than within lineages due to their independent evolution in isolation.

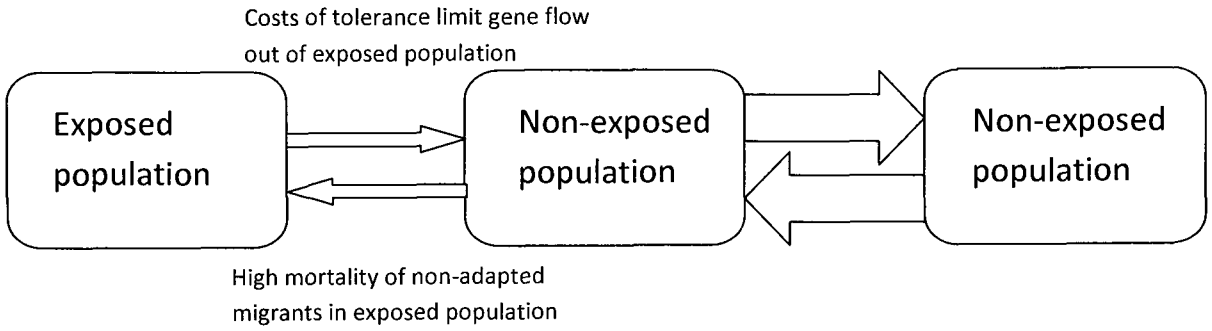
Finally, the third study will apply microsatellite analysis to observe patterns of genetic differentiation between *L. rubellus* populations located along a gradient of nickel contamination. This study will aim to observe whether populations at the most heavily contaminated sites display any characteristics that could indicate that they have undergone selection for metal tolerance. Genetic adaptation will be inferred by observing the degree of genetic differentiation between populations at highly contaminated sites and less contaminated sites. If populations at contaminated sites are genetically adapted, then they may be highly differentiated from the populations at less contaminated sites. The individuals of such populations could be analysed in future exposure-trials to determine if they are more nickel tolerant than individuals from non-

exposed populations. This study will therefore aim to address the following two hypotheses that could account for the presence of *L. rubellus* at nickel-contaminated sites:

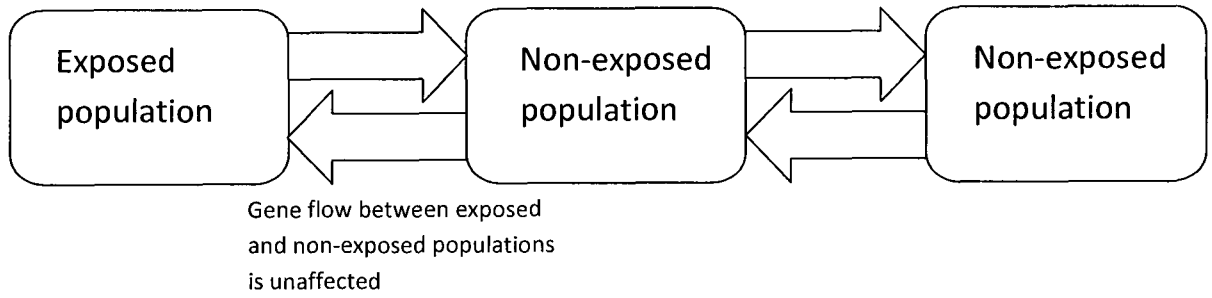
1. The populations are sustained by the local recruitment of genetically adapted nickel-tolerant individuals.

2. The populations represent 'demographic sinks' and are sustained by immigration from less contaminated areas. Individuals within these populations may be able to tolerate the high levels of contamination through a process of phenotypic plasticity.

a)



b)



c)

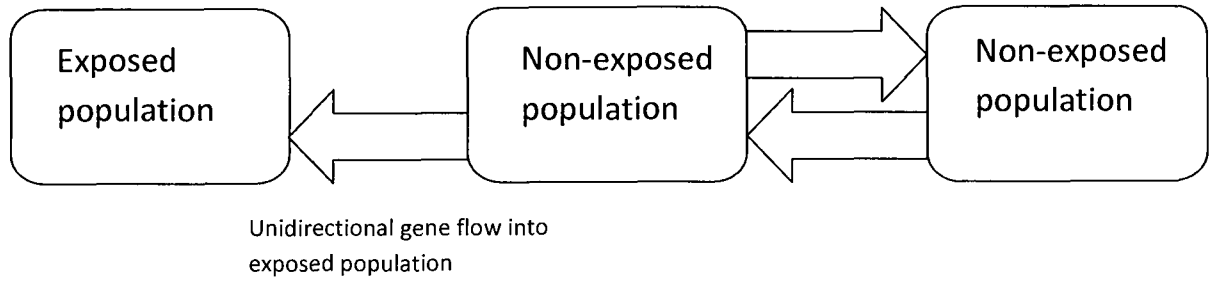


Figure 1.1 Models of expected patterns of gene flow following a) the genetic adaptation of populations at exposed sites, b) the adaptation of populations through phenotypic plasticity, c) the existence of a 'source-sink' dynamic between the populations of non-exposed and exposed sites (arrows indicate both direction and scale of gene flow)

Chapter 2. Intraspecific variation in mtDNA and morphology within the genus *Lumbricus*

The increasing application of molecular techniques to taxonomy has revealed that many species may actually consist of highly divergent 'cryptic' lineages (Bickford et al. 2007). One explanation for this may be that many current species are actually poorly defined and require re-evaluation (Sites and Crandall 1997). In order to increase the accuracy with which species are defined, various concepts have been advanced which typically define species as representing evolutionarily distinct or reproductively isolated lineages (Mallet 2006). Despite the existence of these concepts however, many species remain largely defined on the basis of morphological traits.

The delimitation of species and identification of genetically distinct lineages has however been made easier by the development of DNA barcoding techniques (Hebert et al. 2003, Savolainen et al. 2005, Hajibabaei et al. 2007). DNA barcoding utilises highly conserved DNA sequences such as the mitochondrial cytochrome oxidase I gene (Folmer et al. 1994) that display low levels of intraspecific variation but sufficiently high interspecific variation to allow the resolution of individual species (Hebert et al. 2003). Barcoding genes can also be quickly and reliably sequenced (Hajibabaei et al. 2007), and amplified using general primers that amplify across a wide-range of different taxa (e.g. Folmer et al. 1994). Species can then potentially be identified by alignment of the DNA sequence produced with recognised sequences within a barcode database (Savolainen et al. 2005). Cryptic species can also be potentially identified in barcoding studies if the genetic divergence between them exceeds standard levels of intra-specific variation (Hebert et al. 2003). A further analysis of species resolved by barcoding analysis using behavioural or breeding trials often supports the existence of behavioural or reproductive isolation between them (e.g. Feulner et al. 2006).

The application of DNA barcoding techniques has revealed cryptic variation within many soil-dwelling species (Frati et al. 2004, Heethoff et al. 2006, King et al. 2008, James et al. 2010, Novo et al. 2010b). Such cryptic variation could be a consequence of morphological adaptations to the soil environment, as it has been observed that stressful environments can impose a high degree of stabilising selection (Bickford et al. 2007). Further, many species that are adapted to subterranean environments appear to display a reduction or simplification of external morphological traits (Nevo 1979). The use of morphology to define such species may therefore lead to an underestimation of true species diversity if stabilising selection has imposed constraints upon morphological divergence.

The suitability of DNA barcoding for resolving earthworm taxonomy has been confirmed by several recent studies that have applied the mitochondrial cytochrome oxidase I gene to identify

species, for the construction of phylogenies (Pop et al., 2003; Huang et al., 2007; Pop et al., 2007) or the identification of cryptic species (Chang et al., 2007, James et al. 2010, Novo et al. 2010b). For example, Huang et al. (2007) underlined the accuracy of COI sequence derived taxonomies, showing levels of sequence divergence to be far higher between different earthworm species than between conspecific individuals, confirming the genetic basis of these species groupings. Individual species were also found to successfully group with the other species of their known taxonomic order. The phylogenetic analysis of several European Lumbricid earthworm species using COI primers has recently uncovered evidence of genetically divergent lineages within four different species (*Allolobophora chlorotica, Aporrectodea longa, Aporrectodea rosea* and *Lumbricus rubellus*) (King et al., 2008). Two distinct genetic lineages have been uncovered within *L. rubellus*, displaying a mean sequence divergence of 8.28%. These lineages have been reconfirmed in a recent study analysing amplified fragment length polymorphisms and the mitochondrial cytochrome oxidase II gene (Andre et al., 2010).

The high levels of genetic divergence within *L. rubellus* may therefore be indicative a substantial degree of cryptic genetic heterogeneity within this species. This could have major implications for the role of *L. rubellus* as an effective sentinel species for soil contamination studies. It has been suggested that an important requirement for the selection of effective sentinel species should be that they exhibit genetic homogeneity (Bouché 1992, Hilty and Merenlander 2000). The importance of this requirement has been underlined by studies of cryptic lineages within sentinel species that have revealed differences in their ability to tolerate certain contaminants (Sturmbauer et al., 1999, Linke-Gamenick et al. 2000).

It may however be possible to discriminate between individuals of the two lineages of *L. rubellus* on the basis of morphological differences – this would make practical field work and pollution assessment both reliable and affordable. Several studies of 'cryptic' species have uncovered evidence of previously unrecognised subtle morphological differences upon further analysis (Rocha-Olivares et al. 2001, Wares et al. 2007). In past taxonomic studies of European populations of *L. rubellus*, several putative sub-species have been defined on the basis of minor differences in external morphological traits (Bouché 1972).

In this study the genetic lineages within *L. rubellus* described by King et al. (2008) were further analysed by applying mtDNA to a far larger number of *L. rubellus* individuals. The 'universal' COI primers (Folmer et al., 1994) were applied to investigate the genetic diversity within the two lineages and also to determine whether *L. rubellus* includes further cryptic diversity. Mitochondrial analysis was also applied to individuals of three other Lumbricid species in order to determine whether the lineage diversification within *L. rubellus* is a common feature within this

genus. This is especially relevant, since studies have recently indicated the presence of cryptic genetic variation within the congeneric species, *Lumbricus terrestris* (James et al. 2010).

A morphological study was also conducted to investigate the degree of phenotypic variation within *L. rubellus* and three other *Lumbricus* species. In this study variation in the anterior segment number of the four Lumbricid species was determined, as this particular trait is believed to be evolutionarily conserved and is applied for the taxonomic identification of earthworm species (Sims and Gerard, 1999). Analysis of this trait will also allow identification of possible differences between the two lineages of *L. rubellus*.

A candidate morphological marker, an anteriorally-located glandular tumescence, was also investigated within both lineages of *L. rubellus*. Preliminary observations suggested that this trait may display significant phenotypic divergence between the two lineages and was investigated further in a blind trial to determine if it could be successfully applied to differentiate between the lineages. A lineage-specific mitochondrial Restriction Fragment-Length Polymorphism test was subsequently applied to determine if each inference was correct. Analysis of these traits was carried out to determine whether the lineages within the Lumbricid species may display lineage-specific differences in morphology as has been discovered within some other cryptic species complexes.

Materials and Methods

Sample collection

Specimens of four *Lumbricus* species were collected from a field site located at Pontcanna fields, Cardiff during May 2007 (Table 2.1). Further collections were conducted at two sites at Clydach, Swansea during August 2008. Finally, in order to widen the geographical scope of the study, earthworms were collected from two sites located in Mid-Wales and two sites located in South-West England during August 2010 (See Appendix A for geographical locations). The sample sites represented a range of different habitats including forests, pastureland and park land (See Appendix B for site details). Some of these sites were also contaminated with heavy metals.

Following their collection, all earthworms were washed with distilled water and depurated over 48 hours. During this time the earthworms were kept inside petri-dishes containing two sheets of filter paper dampened with distilled water to prevent desiccation. The filter paper was changed after one day to allow the removal of excreted soil. Following depuration each individual was weighed and then separated into anterior and posterior sections that were then preserved in ethanol. Posterior sections were preserved in 100% ethanol for DNA extraction whilst anterior sections were preserved in 70% ethanol for morphological analysis.

DNA extraction, amplification and sequencing[2]

Approximately 25mg of earthworm tissue was used for each DNA extraction. This tissue was removed from the ethanol-preserved posterior sections using a sterile scalpel blade. DNA was extracted using a QIAGEN DNeasy tissue extraction kit following the protocol for DNA extraction from animal tissue (QIAGEN, UK). A 580bp sub-unit of the mitochondrial cytochrome oxidase I gene (COI) was then PCR-amplified using the general invertebrate primers LCO1490 (5'-GGTCAACAAATCATAAAGATATTGG-3') and HCO2198 (5'-TAAACTTCAGGGTGACCAAAAAATCA-3') (Folmer et al., 1994). PCR reactions were established in 24µl reaction volumes using 2µl of extracted DNA and 0.1µM of each primer added to a PCR mixture containing 1x PCR Buffer (Invitrogen, UK), 3mM Mg2+ (Invitrogen, UK), 0.2mM dNTP mix (dATP, dCTP, dGTP, dTTP) (Invitrogen, UK), 2.5µg bovine serum albumin (New England Biolabs, UK), PCR grade water (Fisher Scientific, UK) and 0.625u Taq polymerase (Invitrogen, UK). All reactions were performed using a GeneAmp PCR System 9700 (Applied Biosystems, UK). All samples were initially denatured for 2.5 minutes at 94ºC and then proceeded through 35 cycles containing a denaturation step at 94ºC for 30s, an annealing step at 47ºC for 30s and an extension step at 72ºC for 45s. After 35 cycles the samples underwent a final extension period of 72ºC for 10 min.

[2] See Fig 2.1 for workflow diagram of mtDNA sequencing

The general COI primers were found to successfully amplify only two of the *L. rubellus* lineage B individuals. A primer pair was therefore developed based upon the COI sequences from these two individuals, which successfully amplified a smaller 356bp fragment of the COI region of the lineage B individuals (LRB COIF 5'-TCTTCTTTCTTGTCATGCCTGT-3', LRB COIR 5'-TGAAGTATTTAGATTTCGGTCAGTT-3').

Reaction mixtures were established as before for the lineage B individuals with 0.1µM of each of the new primers. All reactions were performed using a GeneAmp PCR System 9700 (Applied Biosystems, UK). All samples were initially denatured for 2.5 minutes at 94ºC and then proceeded through 35 cycles containing a denaturation step at 94ºC for 30s, an annealing step at 45ºC for 30s and an extension step at 72ºC for 45s. After 35 cycles the samples underwent a final extension period of 72ºC for 10 min.

After PCR amplification, 10µl of each PCR product was purified using 0.75µl ExoSAP (7.5u Exonuclease I (New England Biolabs, UK) and 0.75u Shrimp Alkaline Phosphotase (Promega, UK)) by incubation at 37ºC for 60 min. Following the reaction, the ExoSAP was inactivated by incubation at 80ºC for 15 min. Amplicons were then sequenced in both forward and reverse orientation. Each reaction contained 2µl of cleaned PCR product, 0.8µM primer, 2µl PCR grade water (Fisher Scientific, UK), 2.5µl Better Buffer (Web Scientific, UK) and 0.5µl Big Dye Terminator v3.1 (Applied Biosystems, UK). All sequencing reactions were carried out following the manufacturers' recommendations (Applied Biosystems, UK). Sequences were visualised using an Applied Biosystems 3130 Genetic Analyzer.

Data analysis

Sequencher 3.1.2. (Gene Codes Corporation, USA) and CodonCode Aligner 3.0.1 (CodonCode Corporation, USA) were used to align and edit all sequences and to determine COI haplotypes for each species. The program MEGA 4.0 (Tamura et al., 2007) was used to determine pairwise genetic distances between all haplotypes. Pairwise distances were calculated both as uncorrected p-distances and also using the Tamura-Nei model of substitution (gamma shape distribution parameter=0.28), selected on the basis of hierarchical likelihood ratio tests performed using the program jModelTest (Posada, 2008). A maximum-likelihood phylogeny was produced using the program PHYML (Guindon & Gascuel, 2003). This was based upon the Tamura-Nei corrected distances between haplotypes with a bootstrap replication of 1000. A Bayesian phylogenetic analysis of haplotypes was also conducted using MrBayes 3.1.2 (Ronquist & Huelsenbeck, 2003). As the Tamura-Nei model of sequence evolution specified by jModelTest is not implemented by MrBayes the general time reversible model was selected with gamma and invariable sites. Bayesian analysis was run using two independent runs consisting of four chains. These analyses

were run for 400, 000 generations with trees saved every 100 generations. Following the analysis, 1000 trees were removed as burn-in. Consensus trees were viewed using TreeView (Page, 1996).

Morphological analysis

Morphological analysis was conducted by counting the anterior segment number between the mouth and clitellum of each specimen under 10x magnification using a light microscope. These segment numbers were then compared with the recognised values for each species (Sims and Gerard, 1999).

The location and size of an anterior genital tumescence was also recorded for all *Lumbricus rubellus* individuals. The genital tumescence was identified under 10x magnification using a light microscope and largely recognised by a conspicuous lack of setae on the segments that it covered.

Images of the genital tumescence were produced using a Jeol LV5200 scanning electron microscope at Cardiff University.

Blind trial

Seventy-five *L. rubellus* individuals were collected from four field sites located throughout South Wales during October 2009. The sites were selected by the collectors so as to include populations of both mitochondrial lineages of *L. rubellus* based upon the findings of previous studies. All earthworms were placed into polythene bags labelled with an identifying code. The code was identifiable only to the collector, allowing a blind trial to be conducted by the experimenter with the collection site of each individual revealed only at the end of the trial.

Following collection all earthworms were processed as described previously. After processing the location and size of the anterior glandular tumescence was recorded for all *L. rubellus* individuals and the possible lineage of the individual recorded based upon the degree of the glandular swelling. The genital tumescence was identified under 10x magnification using a light microscope and largely recognised by a conspicuous lack of setae on the segments that it covered (Figs 2.2 and 2.3).

Sequence analysis of the mtDNA 16s region of lineage A and B *L. rubellus* yields two restriction sites that are present within lineage B but absent from lineage A (A. King, pers. comm.). By applying restriction fragment analysis to PCR amplified fragments, it was therefore possible to quickly and efficiently determine the lineage of *L. rubellus* individuals.

DNA was extracted from the posterior section of each sample using a QIAGEN DNeasy tissue extraction kit (QIAGEN, UK). A 500-650 bp fragment of the mtDNA 16s RNA coding region was then PCR amplified in a 25µl reaction using the general invertebrate primer pair 16sar (5'-

CGCCTGTTTATCAAAAACAT-3') and 16sbr (5'-CCGGTCTGAACTCAGATCACGT-3') (Palumbi et al., 1991). The PCR reactions were set up using 1µl of the extracted DNA and 0.5µM of each primer added to a PCR mixture containing 1x PCR Buffer (Invitrogen, UK), 4mM Mg2+ (Invitrogen, UK), 0.4mM dNTP mix (dATP, dCTP, dGTP, dTTP) (Invitrogen, UK), PCR grade water (Fisher Scientific, UK) and 0.625u Taq polymerase (Invitrogen, UK). All reactions were performed using a GeneAmp PCR System 9700 (Applied Biosystems, UK). All samples were initially denatured for 2.5 minutes at 94ºC and then proceeded through 35 cycles containing a denaturation step at 94ºC for 30s, an annealing step at 47ºC for 30s and an extension step at 72ºC for 45s. After 35 cycles the samples underwent a final extension period of 72ºC for 10 min. Restriction digestion of the 16s fragments was then performed using the restriction enzymes DraI and HpyI in 10µl reaction volumes containing 7.8µl of the PCR product, 1× enzyme buffer (NE Biolabs Ltd., UK), distilled water and 4 units of restriction enzyme (NE Biolabs Ltd., UK). All samples were then incubated at 37ºC for 3h using a GeneAmp PCR System 9700 (Applied Biosystems, UK). Following incubation all fragments were then run on a 1.5% agarose gel.

Table 2.1 Numbers of sampled individuals at each site

Collection site	Site code	*L. castaneus*	*L. festivus*	*L. rubellus*	*L. terrestris*
Pontcanna fields, Cardiff	PT	17	25	25	22
Clydach (site 1), Swansea	CL1	0	0	22	0
Clydach (site 2), Swansea	CL2	0	11	22	9
Gloversfield, Somerset	GF	0	0	19	0
Hallen Hill, Bristol	HH	0	0	9	0
Ystwth Source, Powys	YS	0	0	10	0
East Cottage, Ceredigion	EC	0	0	16	0
Wemyss, Ceredigion	WM	0	0	14	0
Total		17	36	137	31

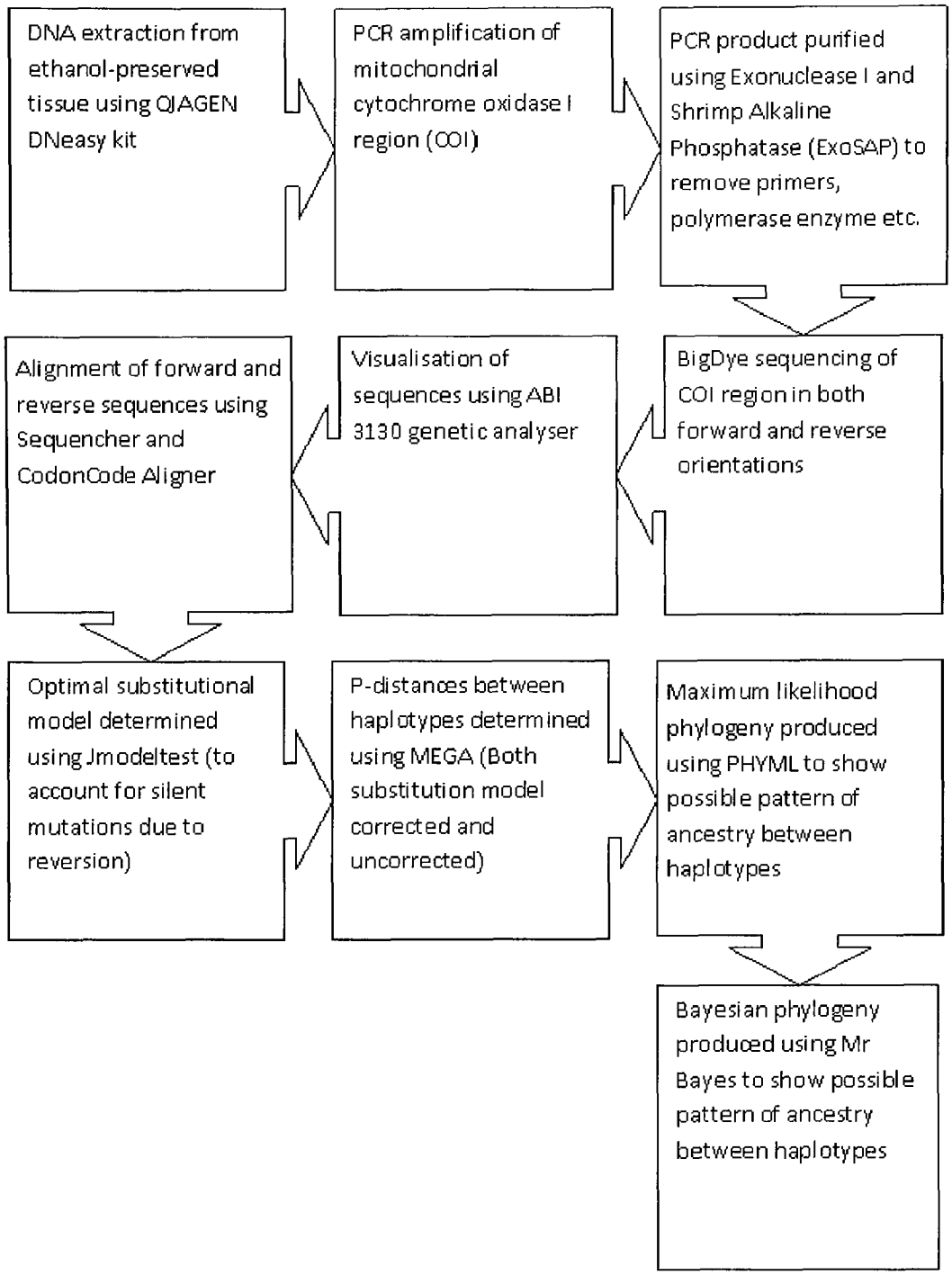


Figure 2.1 Work-flow diagram of mtDNA sequence analysis methodology

Results

Molecular analysis

A 356 bp sub-unit of the mitochondrial COI gene between base positions 238 and 593 was successfully sequenced for 221 earthworms in total. These sequences were found to align unambiguously and displayed no evidence of insertion/deletions.

Twenty-two COI haplotypes were found within *L. rubellus*. The haplotypes were found to group into two distinct clades corresponding to the previously described lineage A and B of *L. rubellus* (King et al., 2008). Nine haplotypes were uncovered among the 'lineage A' *L. rubellus* (Table 2.1). Thirteen haplotypes were uncovered among the 'lineage B' *L. rubellus*.

Of the lineage A haplotypes, four were found only in South Wales (rubellus 3, 4, 5 and 6) (Table 2.3). Of the remaining haplotypes, rubellus 1 was found in South Wales and South-West England, rubellus 7 was found in North Wales and South Wales, and rubellus 8 and 9 were found in North Wales and South-West England. Rubellus 1 was the only haplotype to be found in all regions.

Within lineage B, five haplotypes were found only in North Wales (rubellus 13, 14, 15, 16 and 22), two haplotypes were found only in South Wales (rubellus 10 and 12) and five haplotypes were found only in South-West England (rubellus 17, 18, 19, 20 and 21). Rubellus 11 was found in both South Wales and South-West England.

Within both lineages no clear correlation could be discerned between either the habitat type or contamination status of a site and its constituent haplotypes. In lineage A, four of the haplotypes found at the non-contaminated site of Pontcanna were not found at any of the other sample sites. However this may be a result of the greater sample size of this site compared with the other sites.

Within the other three species, eight COI haplotypes were found within *Lumbricus castaneus*, seven haplotypes were found within *Lumbricus festivus* and seven haplotypes were found within *Lumbricus terrestris*. Of the seven *L. terrestris* haplotypes, three were found only at the Pontcanna site (Terrestris2, 3 and 4) and another three were found only at the Clydach site (Terrestris5, 6 and 7). The remaining haplotype, Terrestris1, was found at both sites and was particularly abundant at Pontcanna where it was expressed by 82% of the sampled individuals. Of the seven *L. festivus* haplotypes, two were found only at the Pontcanna site. The other five haplotypes were found at both sites.

The mean uncorrected p-distance of 0.14 between the lineage A and B *L. rubellus* was found to slightly higher than that previously described for *L. rubellus* by King et al. (2008) and is almost comparable to the distances observed between *L. rubellus* and the other *Lumbricus* species

featured in this study (Table 2.2). Genetic distances between haplotypes within *L. rubellus* lineage B were much lower than between those within lineage A with a mean uncorrected p-distance of 0.01. The haplotypes of lineage A showed a mean uncorrected p-distance of 0.02 which was comparable to genetic distances between the haplotypes of *L. festivus* (Table 2.2) but lower than the intraspecific genetic distances observed between haplotypes of *L. castaneus* and *L. terrestris*.

Both the ML and Bayesian phylogenies displayed strong support for the monophyly of all of the *Lumbricus* species analysed (Figs 2.4 and 2.5). The phylogenies also both showed strong support for the monophyly of both lineage "a" *L. rubellus* and lineage "b" *L. rubellus*. The bootstrap and posterior probabilities of the two phylogeneies also supported the existence of haplogroups within *L. rubellus* lineage "a", *L. castaneus*, *L. festivus* and *L. terrestris*.

Morphological analysis

Morphological analysis of anterior segment number revealed this particular morphological character to be highly conserved in all four species. Within *L. rubellus* all lineage B individuals were found to display the described segment number of 26 and only a single lineage A individual was found to display a segment number which differed from this value. Within *L. terrestris*, only one individual was found to differ from the described segment number of 31, by displaying 29 segments and in *L. festivus*, two individuals were found to differ from the described segment number of 33, by displaying 32 segments. *L. castaneus* displayed the highest level of morphological variation with three individuals differing in anterior segment number from the described segment number of 27 by displaying 26 segments.

Morphological analysis of *L. rubellus* also uncovered a polymorphic trait within both lineage A and B individuals. A genital tumescence was discovered in all adult individuals associated with a lack of ventral setae (Figs 2.2 and 2.3). The location of this swelling was found between segments 8 and 12 in lineage A and segments 9 and 12 in lineage B. The tumescence was only found on a single segment for all lineage B individuals whereas the large majority of lineage A individuals displayed the swelling over two segments. It was also noted that the genital tumescence appeared more easily identifiable within lineage B individuals and was often observable without magnification.

Blind trial

The blind trial experiment to identify *L. rubellus* lineage A and B individuals on the basis of differences in the expression of a genital tumescence led to the successful identification of 80% of the individuals. This was found to be statistically significant (chi^2 test, $p=<0.05$). Observation of the observed and expected values indicate that 89% of lineage B individuals were correctly identified which was higher than the expected value of 79%. The correct identification of lineage

A individuals was largely unsuccessful though with 71% of individuals successfully identified which was lower than the expected value of 79%.

Table 2.2 Mean pairwise differences between *Lumbricus* species and lineages (Tamura-Nei distance above diagonal, non-corrected p distance below diagonal), with mean intraspecific values along the diagonal (non-corrected p distance/Tamura-Nei distance)

	1	2	3	4	5
1. *L. rubellus* (lineage a)	0.02/0.03	0.30	0.56	0.37	0.17
2. *L. rubellus* (lineage b)	0.14	0.01/0.01	0.20	0.18	0.15
3. *L. castaneus*	0.21	0.50	0.05/0.06	0.22	0.20
4. *L. festivus*	0.17	0.38	0.60	0.02/0.03	0.19
5. *L. terrestris*	0.39	0.30	0.49	0.46	0.05/0.06

Table 2.3. Haplotype frequency at different sites

	CR1	CR2	EC	GF	HH	PT	WM	YST
L. castaneus 1						4		
L. castaneus 2						4		
L. castaneus 3						2		
L. castaneus 4						1		
L. castaneus 5						3		
L. castaneus 6						1		
L. castaneus 7						1		
L. castaneus 8						1		
L. festivus 1	4					15		
L. festivus 2						1		
L. festivus 3	1					2		
L. festivus 4	2					3		
L. festivus 5	1					2		
L. festivus 6						1		
L. festivus 7	3					1		
L. rubellus 1	9		3	1	6	15	6	2
L. rubellus 2	8				1	3		
L. rubellus 3						1		
L. rubellus 4						1		
L. rubellus 5						1		
L. rubellus 6						2		
L. rubellus 7	5		1			2		
L. rubellus 8			1		1		3	
L. rubellus 9			1		1			
L. rubellus 10		8						
L. rubellus 11		12		7				
L. rubellus 12		2						
L. rubellus 13			6					7
L. rubellus 14							1	
L. rubellus 15							2	
L. rubellus 16								1
L. rubellus 17				1				
L. rubellus 18				1				
L. rubellus 19				1				
L. rubellus 20				1				
L. rubellus 21				2				
L. rubellus 22			1					
L. terrestris 1	3					18		
L. terrestris 2						1		
L. terrestris 3						2		
L. terrestris 4						1		
L. terrestris 5	2							
L. terrestris 6	3							
L. terrestris 7	1							

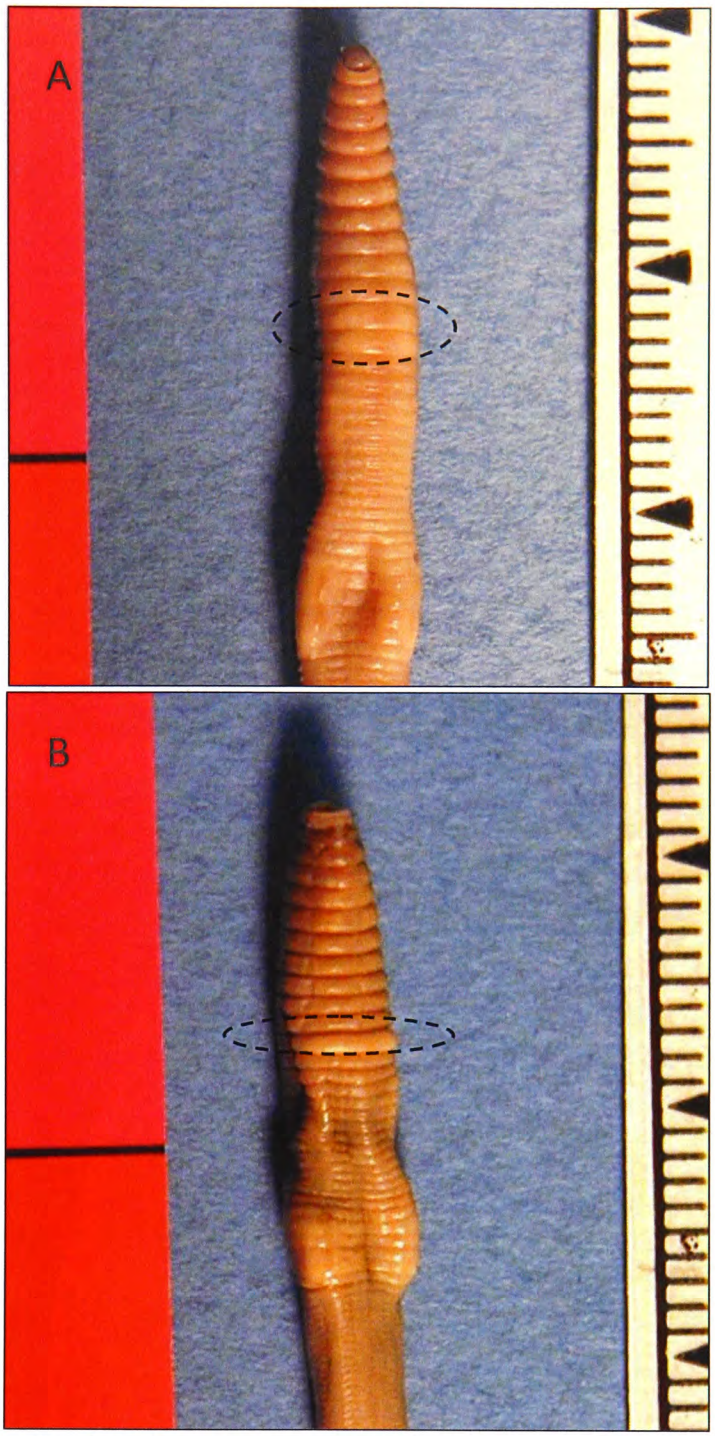


Figure 2.2 10x magnification of anterior sections of A) *L. rubellus* lineage A and B) *L. rubellus* lineage B (Glandular tumescence is circled)

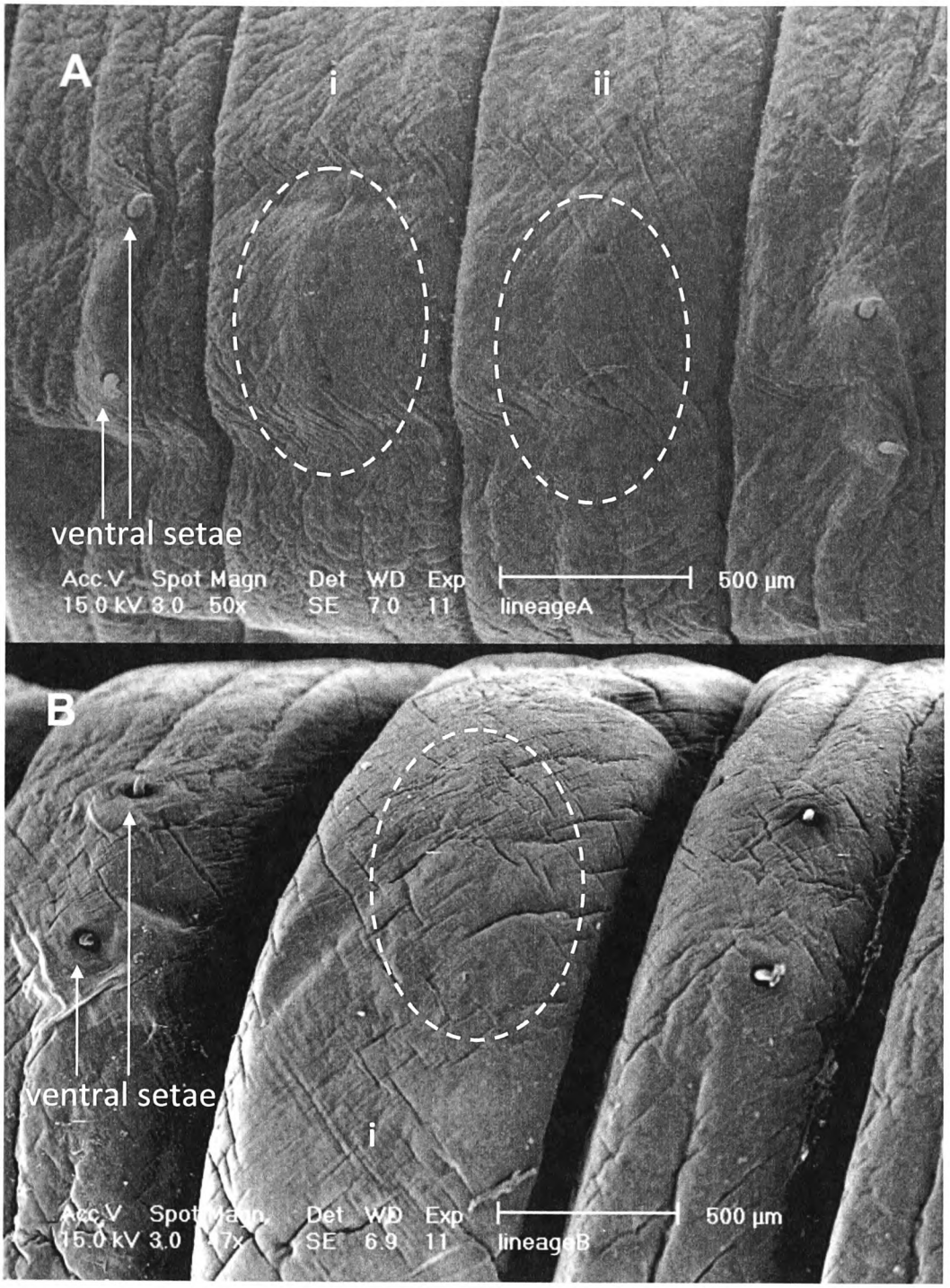


Figure 2.3 Scanning electron microscope images of glandular tumescences of A) *L. rubellus* lineage A (segments i-ii) B) *L. rubellus* lineage B (segment i). Note the absence of ventral setae on these segments.

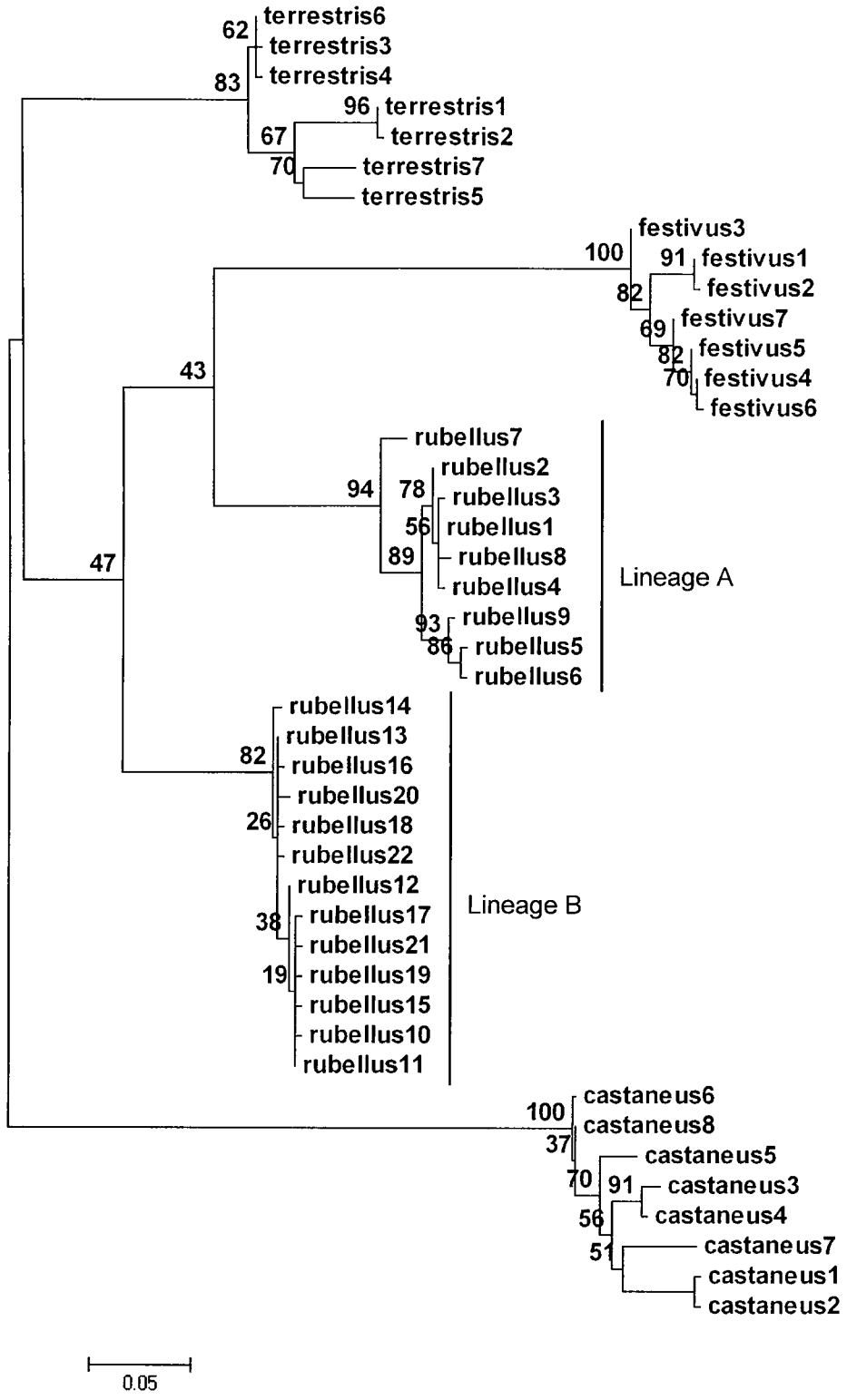


Figure 2.4 Maximum Likelihood tree of *Lumbricus* mtDNA COI haplotypes. Genetic distances calculated using the Tamura-Nei model of substitution. Numbers on tree indicate bootstrap support obtained with 1000 replicates.

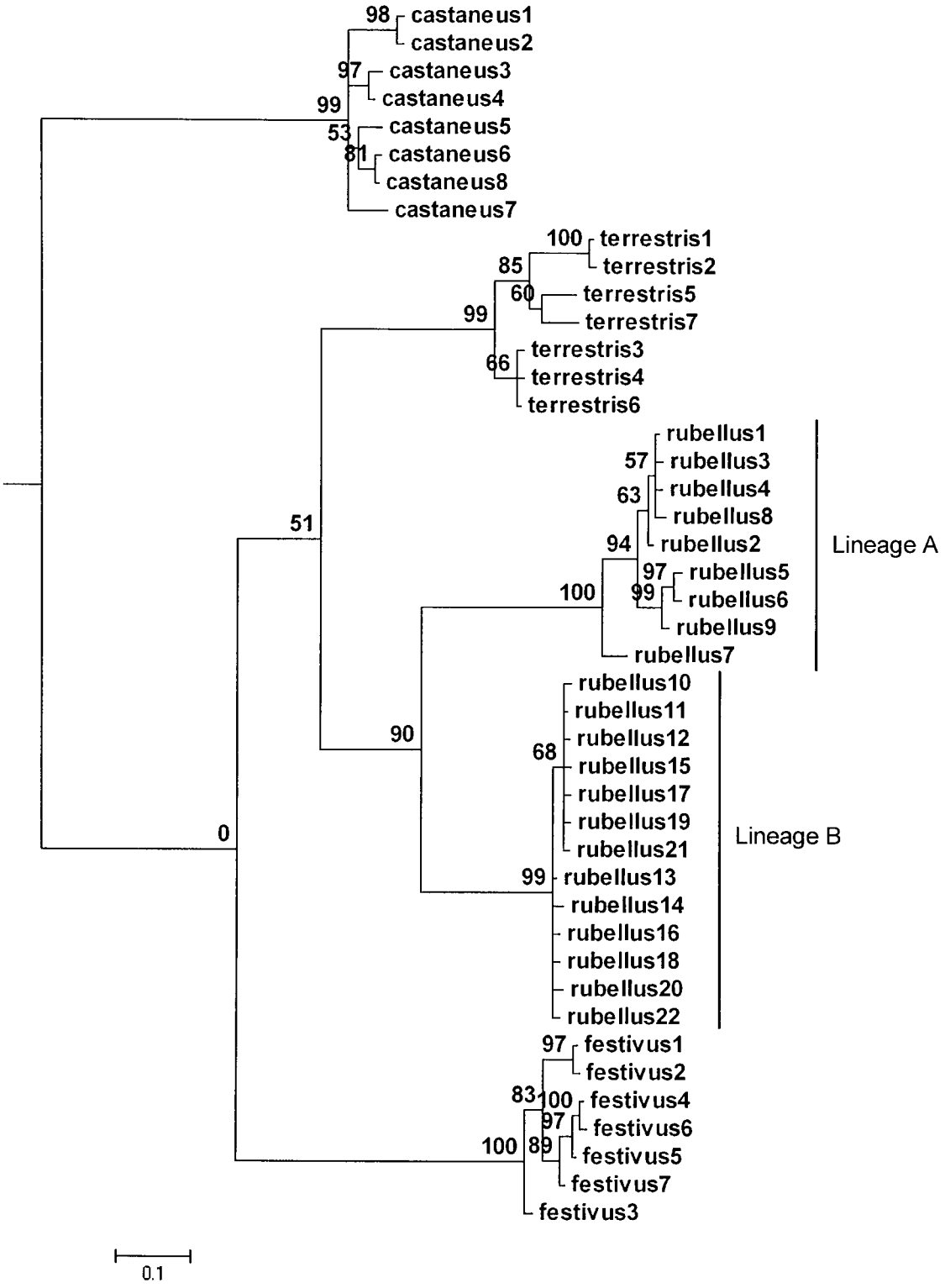


Figure 2.5 Majority-rule consensus tree from the Bayesian analysis of *Lumbricus* mtDNA COI haplotypes. Genetic distances calculated using the GTR+I+Γ model of substitution. Numbers on tree indicate posterior probabilities.

Discussion

Analysis of genetic variation in the mtDNA COI gene of *L. rubellus* confirmed the existence of the two cryptic mitochondrial lineages uncovered by King et al. (2008). The mean divergence between the haplotypes of lineage A and B was relatively high (uncorrected p-distance = 0.14) and was comparable to the level of divergence found between *L. rubellus* and the three congeneric species featured in the study (uncorrected p-distances = 0.17-0.39).

Although none of the other three species analysed were found to contain haplogroups comparably divergent to those found within *L. rubellus*, the average levels of genetic variation were still found to be relatively high compared with general values from COI barcoding studies of earthworms, which show levels of intraspecific variation to be largely below 1% (Huang et al., 2007). Several well-supported lineages were found within *L. castaneus* and *L. terrestris*, and the *L. rubellus* lineage A haplotype, rubellus 7 was found to diverge from the other haplotypes by p-distances of between 0.01 and 0.05.

The fact that both lineages of *Lumbricus rubellus* were found at nickel contaminated sites (Clydach) indicates that they could pose a problem for ecotoxicological studies applying *L. rubellus* as a sentinel species. In order for ecotoxicological studies to yield accurate results about the effects of contaminants, it has been suggested that the earthworm species used within them should be selected on the basis of being genetically homogeneous (Bouché, 1992). The importance of this assertion is underlined by a study of the aquatic oligochaete species *Tubifex tubifex*, which showed the individuals of different lineages diverging in 16S rDNA by up to 13%, to display differences in their susceptibility to cadmium contamination (Sturmbauer et al., 1999). It is therefore possible that the two lineages of *L. rubellus* could display differences in their levels of metal tolerance at a contaminated site.

One way in which the possible confounding effects of the two *L. rubellus* lineages could be overcome in ecotoxicological studies may be through the application of lineage specific morphological traits during sample collection. The degree of variation in two morphological traits; anterior segment number and an anterior glandular tumescence, was therefore analysed to determine whether any differences could be observed between the two lineages. Morphological analysis revealed anterior segment number to be a largely well conserved trait both within the two *L. rubellus* lineages and the three other *Lumbricus* species, with only a very small proportion of individuals showing any variation from the anterior segment numbers described by Sims and Gerard (1999). All variations were small and did not vary from the described values by more than three segments. All individuals displaying variation were also found to display a segment number that was lower than the described values. The largely highly conserved nature of anterior segment number within *Lumbricus* species may be due to several different factors. Firstly it has

been proposed that organisms surviving in highly stressful conditions such as those that are found in subterranean environments may face strong stabilising selection upon their morphology that constrains any major changes in body structure (Lefébure et al., 2006, Bickford et al., 2007). Previous morphological studies have also indicated that anterior segment number may be closely correlated to the ecological niche of earthworm species enabling differences in locomotory behaviour between burrowing anecic species and litter-dwelling epigeic species (Piearce, 1983). Secondly, as the mechanism of copulation in *Lumbricus* involves individuals joining up 'head-to-tail' (Edwards and Bohlen, 1996), any major differences in anterior segment number might result in a loss of fitness as reproductive structures could fail to align correctly during mating. This could therefore also impose stabilising selection upon anterior segment number.

Contrastingly, morphological analysis of the anteriorally located glandular tumescence revealed a wide degree of variation in both its location and within lineage A, in its size indicating it to be a morphologically plastic trait. The swelling appeared to be far more pronounced and stable within individuals of lineage B which suggested its possible utility in allowing the two lineages to be successfully sorted during field collection. The findings of the blind trial indicated that >80% of the *L. rubellus* individuals were correctly identified on the basis of morphological differences in this trait – and nearly 90% of lineage B individuals. Statistical analysis revealed this to be significantly different to the expected success rate for lineage B. It is therefore possible that the application of more detailed methods of morphological identification during field collection such as the use of portable digital microscopes, could lead to an effective assay in using this trait to differentiate the two lineages. An accurate means of correctly identifying lineage A individuals was also suggested by the fact that swellings spanning two segments or more were found to be unique to this particular lineage. The collection of *L. rubellus* individuals displaying this characteristic could therefore be used to ensure that all individuals are of the same lineage

The glandular tumescence observed within both lineages of *L. rubellus* is relatively poorly described compared to other morphological traits such as the clitellum (Sims and Gerard, 1999). Differences in the size, segment number and extent of this trait have however been used to characterise two possible taxonomic sub-species of *L. rubellus* (Bouché, 1972).

Given that the two distinct lineages were uncovered using maternally-inherited mitochondrial markers, there does however remain a possibility that male-mediated introgression could be occurring between the two lineages. The AFLP genotyping of some *L. rubellus* populations has indicated the possible existence of hybrid individuals (Andre et al., 2010). A further nuclear-gene based phylogenetic analysis of the two lineages is therefore required to determine whether these two lineages are reproductively isolated.

It has been proposed that the genetic heterogeneity uncovered within this study may be a common feature of earthworm species (Stürzenbaum, 2009). The application of mitochondrial analysis techniques to studies of earthworm phylogenetics has already uncovered evidence of substantial cryptic variation within many earthworm species (Heethoff et al., 2004, Huang et al., 2007, King et al., 2008, Chang et al., 2009; Pérez-losada et al., 2009; Andre et al., 2010). One possible explanation for the existence of a high degree of genetic divergence between the cryptic lineages of some British earthworm species, is that they represent divergent lineages which inhabited different glacial refugia during the late Pleistocene (King et al., 2008). Many animal and plant species within Northern Europe display evidence of genetic lineages which are believed to represent populations drawn from different South European refugia (Taberlet et al., 1998). However, the majority of these lineages occur allopatrically, unlike the two sympatric lineages of *L. rubellus*.

An explanation for the existence of sympatric lineages could be that some earthworm lineages may have developed from populations that survived the late Pleistocene within periglacial cryptic refugia close to their present day distributions (Stewart and Lister, 2001). These lineages could have merged with populations from the southern refugia during the period of post-glacial re-colonisation. The survival of small isolated populations might have resulted in population bottlenecks explaining the low levels of genetic diversity that have been observed within some cryptic lineages of earthworms (King et al., 2008). The survival of earthworms in periglacial refugia is supported by the discovery of contemporary troglophilic populations of several lumbricid earthworm species including *L. rubellus* (Reeves, 1999). Periglacial survival has been proposed as accounting for the existence of populations of the cold-tolerant earthworm *Dendrobaena octaedra* in Greenland (Hansen et al., 2006).

However an alternative hypothesis may be that *L. rubellus* diversified following its re-colonisation of Britain. The glaciations of the Pleistocene are believed to have largely eradicated the earthworm fauna of Britain (King et al. 2008). Although species such as *L. rubellus* were able to successfully recolonise Britain, the formation of the English Channel prevented the arrival of many other earthworm species (King et al. 2008). This resulted in Britain displaying a lower overall species diversity than the Northern European mainland (Sims and Gerard 1999). This could have left many ecological niches open for the diversification of re-colonising earthworm species such as *L. rubellus*. Such a process of diversification has been proposed in other studies where species are believed to have colonised environments with vacant ecological niches (Dimitrov et al. 2008, Burbrink and Pyron 2009).

In conclusion molecular analysis confirmed the existence of two distinct lineages within *L. rubellus* at several sites throughout Great Britain and uncovered evidence of slightly high levels of

genetic heterogeneity within *L. rubellus* lineage A and some of the other *Lumbricus* species (*L. castaneus* and *L. terrestris*). The two lineages of *L. rubellus* could pose a problem for ecotoxicologists applying the species as a ecological sentinel, as they could theoretically respond differently to contaminants (Bouché, 1992) as has been found for other oligochaete species displaying cryptic variation (Sturmbauer, 1999). Morphological analysis of the two lineages uncovered a trait that could possibly be used to differentiate between them. A blind morphological trial indicated lineage-sorting on the basis of this trait to be significantly effective in differentiating the lineages.

Chapter 3. A comparison of microsatellite variation in cryptic lineages of *Lumbricus rubellus*

In recent years mtDNA sequence analysis has uncovered evidence of substantial cryptic variation within a range of earthworm species (King et al. 2008, Andre et al. 2010, James et al. 2010, Novo et al. 2010b). Such cryptic variation could pose a problem for the use of earthworm species within studies of soil ecotoxicology as one of the key requirements for ecotoxicological sentinel species is that they display genetic homogeneity (Bouché 1992, Hilty and Merenlander 2000). Exposure trials of various contaminants have revealed that cryptic lineages within some sentinel species may differ in tolerance, resulting in different frequencies of these lineages in contaminated and reference populations (Sturmbauer et al. 1999, Rocha-Olivares et al. 2004). Given these potential differences in tolerance, it is important that cryptic lineages are studied in species utilised in ecotoxicology, as inferences of ecosystem health based upon cryptic species could be misleading given that one of the key requirements of an ecotoxicological sentinel species is that it displays genetic homogeneity (Bouché 1992, Hilty and Merenlander 2000).

The cosmopolitan epigeic earthworm species, *Lumbricus rubellus* which is widely applied as a sentinel species within the field of soil ecotoxicology, has been found to comprise two distinct mtDNA lineages (King et al. 2008, Andre et al. 2010). A multi-locus nuclear DNA based analysis using AFLP markers was also able to differentiate individuals of the two lineages, although principal component analysis indicated the possibility of some hybridisation between them (Andre et al. 2010). Differences in heavy metal tolerance between the lineages of *L. rubellus* may be possible when it is considered that many earthworm species appear to differ in their tolerance of contaminants, indicated by differences in the earthworm species compositions of contaminated and reference sites (Lukkari et al. 2004). Such differences may be a consequence of species-specific variation in metal accumulation, which have been found to exist between earthworm species (Morgan and Morgan 1991, Dai et al. 2004, Kamitani and Kaneko 2007) including species operating within a similar ecological niche (Morgan and Morgan 1991).

Sympatric populations of the two lineages of *L. rubellus* may also differ in tolerance due to differences in their capacity to adapt to environmental stress. It has been suggested that the populations of even a single species could vary in this respect due to differences in genetic variation (Morgan et al. 2007). The study of Andre et al. (2010) also provides circumstantial evidence for differences in metal tolerance between the two lineages of *L. rubellus*. The study uncovered differences in the distribution of the two lineages across a site heterogeneously contaminated with lead. One lineage was found to be abundant at a heavily contaminated site, whereas the other lineage was found to be abundant at a nearby calcareous site with lower levels of lead contamination.

In this study, microsatellite markers were applied to analyse genetic diversity within and differentiation between the two lineages within *L. rubellus*. To observe whether there is any evidence of introgression between the two lineages, microsatellite allele frequencies were analysed in sympatric populations of the two lineages at sites in South Wales, UK. Many studies have successfully applied microsatellite frequency data to demonstrate the existence of reproductive isolation, often applying the same statistical approaches as in population genetics. This includes the determination of population differentiation through the calculation of coefficients such as F-statistics to infer the degree of gene-flow between populations. This approach has successfully demonstrated the existence of reproductive isolation within several cryptic species complexes (Feulner et al. 2006, Racey et al. 2007, Li et al. 2009). Interspecific studies have also applied exploratory data analysis such as Bayesian clustering to observe the degree of demographic isolation between sympatric populations of different species. This approach has been particularly useful for studies aiming to uncover evidence of introgression between cryptic lineages (Luo et al. 2010, Väli et al. 2010, Zeng et al. 2010).

Several studies of terrestrial invertebrates have applied microsatellite allele frequencies to determine the degree of genetic differentiation between species or cryptic lineages. Streiff et al. (2005) applied microsatellite analysis to determine the extent of hybridisation between two sympatric species of carabid beetle. An understanding of the degree of introgression between these species was necessary due to their widespread use in biomonitoring studies. Microsatellite analysis of a number of populations indicated a high degree of differentiation between the two species. Bayesian techniques did however indicate the existence of possible back-cross hybrids at three sites. It was therefore proposed that hybridisation between the two species may be occurring at a low level. Secondly, the study of Roy et al (2006) applied microsatellite analysis in combination with nuclear and mtDNA sequence analysis to determine the degree of differentiation between populations of a soil-feeding termite. The population under analysis was suspected to actually comprise several cryptic species, with a lack of morphological variation resulting in the taxa being poorly described by taxonomists. The results of the microsatellite analysis were in agreement with the sequence data, revealing a high degree of genetic differentiation between colonies believed to represent different cryptic species. Finally, a recent microsatellite analysis of the earthworm species *Allolobophora chlorotica* applied factorial correspondent analysis to investigate several cryptic mitochondrial lineages that have been discovered within the species (Dupont et al. 2011). The FCA analysis displayed separate clusters for several of the lineages. However, the location of some individuals within clusters of a different lineage was however seen as indicating the possible existence of hybrid individuals.

Microsatellite loci have also been sequenced in this study to enable the identification of lineages-specific motifs within microsatellite flanking (MFR) and repeat regions. Several studies have

demonstrated that the sequencing of microsatellite loci represents an effective alternative to allele frequency analysis for studies of interspecific differentiation (Fisher et al. 2000, Makova et al. 2000, Feulner et al. 2006). One advantage of this approach is that problems resulting from size homoplasy that can confound other microsatellite-based phylogenetic approaches can be overcome. MFR sequences have already been successfully applied to derive phylogenies for cryptic species complexes (Fisher et al. 2000, Feulner et al. 2006). Interspecific comparisons of microsatellite flanking sequences have also indicated the existence of sequence differences between closely related species (Makova et al. 2000). Levels of variation within these species were found to be comparatively low. Microsatellite flanking regions therefore offer a similar capacity for the resolution of individual species to the nuclear and mtDNA sequences that are currently applied in genetic barcoding.

In this study two alternative methods of microsatellite analysis were therefore applied to investigate the degree of nuclear genetic differentiation that exists between the two cryptic lineages of *L. rubellus*. This would enable the identification of possible male-mediated introgression between the lineages[3]. Such male-mediated intogression can often be overlooked by molecular studies given the widespread application of mitochondrial markers for genetic bar-coding (Moritz and Cicero 2004). This can lead to the existence of some incongruence between the findings of mtDNA studies and other nuclear-DNA based studies. In order to gain a greater understanding of the cryptic lineages within *L. rubellus* it is therefore important that nuclear genetic differentiation is assessed.

The following hypotheses will be tested in this study-

1. Lineage A and B are reproductively isolated with a clear differentiation between individuals of the two lineages. Genetic differentiation between sympatric populations of the two lineages is higher than that between allopatric populations of a single lineage
2. Lineage A and B are capable of genetic hybridisation with some individuals displaying hybrid genotypes

[3] It has been suggested that the paternal inheritance of mitochondria could occur within some hermaphroditic animals that are able to produce male and female gametes in a single organ (Davison 2006). However this should not occur within *L. rubellus* as although the species is a simultaneous hermaphrodite, the male and female gametes are produced within separate organs (Edwards and Bohlen 1996).

3. Microsatellites display a greater degree of divergence in both sequences and repeat units between the lineages than within lineages. This results from their independent evolution following the development of reproductive isolation.

Materials and methods

Microsatellite sequencing[4]

PCR amplification, purification and cloning

Eight microsatellites were PCR amplified in *L. rubellus* individuals of both cryptic mtDNA lineages (Table 3.4). For each microsatellite, PCR amplification was conducted using ten 'lineage A' individuals and ten 'lineage B' individuals from chapter 1. All lineage A individuals were selected from the Pontcanna and Clydach 1 populations (see Appendix B for site details). All lineage B individuals were selected from the Clydach 8 population.

PCR reactions were set up in 20µl reaction volumes using 1µl of extracted DNA and 0.2µM of each primer added to a PCR mixture containing 1x PCR Buffer (Invitrogen, UK), 3mM Mg2+ (Invitrogen, UK), 0.2mM dNTP mix (dATP, dCTP, dGTP, dTTP) (Invitrogen, UK), PCR grade water (Fisher Scientific, UK) and 1u Taq polymerase (Invitrogen, UK). All reactions were performed using a GeneAmp PCR System 9700 (Applied Biosystems, UK). All samples were initially denatured for 5 minutes at 94ºC and then 35 cycles containing a denaturation step at 94ºC for 1 minute, an annealing step at a primer-specific temperature (see Table 3.4) for 1 minute and an extension step at 72ºC for 1 minute was carried out. After 35 cycles the samples underwent a final extension period of 72ºC for 5 min.

Microsatellite DNA was purified by the gel electrophoresis of 20µl of each PCR reaction on a 3% agarose gel stained with ethidium bromide. DNA was visualised by illumination with a U.V. light box and microsatellite DNA identified by reference to a 100b.p. ladder. Microsatellite DNA was removed from the gel by cutting out the illuminated band using a sterile razor blade. The excised gel was then placed into a 1.5ml eppendorf tube containing 10µl PCR grade water (Fisher Scientific, UK). Using a pair of tweezers the gel was positioned so that only a small part of it was in contact with the water. Each sample was then left for a period of approximately 12h to allow the DNA to diffuse into the water. The gel fragment was then removed and each DNA sample stored at -20°C.

Four libraries were produced by TA cloning (Table 3.2). Each library was constructed using five different DNA fragments for each of the microsatellite loci in that particular library. Microsatellite cloning reactions were performed using an AccepTor Vector TA cloning kit (Novagen).

The pooled gel-purified microsatellite fragments were ligated into linearised pSTBlue-1 cloning vectors in 10µl reactions that featured 50ng of acceptor vector, 0.5µl of ~10ng/µl purified DNA

[4] See Fig 3.1 for workflow diagram of microsatellite sequencing

fragments, 1X Clonables Ligation Premix and nuclease-free water. The reaction was left to ligate at room temperature for twenty minutes.

Following ligation, 1µl of the reaction was used to transfect an aliquot of NovaBlue Singles Competent Cells. Each aliquot tube was placed on ice for 5 min and then placed into a water bath set at 42°C for 30 sec to allow heat shock transformation of the cells. Transformation was terminated by placing each tube back onto ice for two minutes. Prior to plating, an outgrowth step was performed to increase cell quantity with 250µl of room temperature SOC medium added to each tube. The tubes were placed on an orbital shaker (125 rpm) inside an incubator at 37°C for a period of 1 hour.

The transformed cells were then spread onto four petri-dishes of LB agar (Sigma-Aldrich), each containing 15µl of 50mg/ml kanamycin to remove the non-transformed cells. One of these plates was spread with 100µl of cells combined with 15µl SOC medium. The other three plates were spread with 50µl of cells combined with 20µl SOC medium. Each petri-dish was pre-prepared with 87.5µl of 20mg/ml X-gal and 20µl of 100mM IPTG to allow blue-white screening. Plates were incubated at 37°C for 12h.

Following incubation, bacterial colonies containing fragments were uncovered by blue-white screening. Each successfully transformed colony was picked with a sterile pipette tip and used to inoculate a single well on a 96-well microplate (Corning) containing 200µl LB broth with 0.12µl 50mg/ml kanamycin to allow further growth of the colony. All plates were placed on an orbital shaker (70 rpm) inside an incubator at 37°C for a period of 12 hours.

PCR amplification and sequencing of insert DNA

Insert fragments were PCR amplified directly from transformed clone colonies using the insert primers T7 (5'-TAATACGACTCACTATAGGG-3') and SP6 (5'-ATTTAGGTGACACTATAGA-3'). All PCR reactions were set up in 10µl reaction volumes using 0.5µl of the clone colony and 0.1µM of each primer added to a PCR mixture containing 1x PCR Buffer (Invitrogen, UK), 3mM Mg2+ (Invitrogen, UK), 0.2mM dNTP mix (dATP, dCTP, dGTP, dTTP) (Invitrogen, UK), PCR grade water (Fisher Scientific, UK) and 1u Taq polymerase (Invitrogen, UK). All reactions were performed using a GeneAmp PCR System 9700 (Applied Biosystems, UK). All samples were initially denatured for 5 minutes at 94ºC and then proceeded through 35 cycles containing a denaturation step at 94ºC for 1 minute, an annealing step at 50ºC for 1 minute and an extension step at 72ºC for 1 minute. After 35 cycles the samples underwent a final extension period of 72ºC for 5 min.

After PCR amplification, 10µl of each PCR product was purified using 0.75µl ExoSAP (7.5u Exonuclease I (New England Biolabs, UK) and 0.75u Shrimp Alkaline Phosphotase (Promega, UK))

by incubation at 37ºC for 60 min. Following the reaction, the ExoSAP was inactivated by incubation at 80ºC for 15 min. Amplicons were then sequenced in both forward and reverse orientation. Each sequencing reaction contained 2µl of cleaned PCR product, 0.8µM primer, 2µl PCR grade water (Fisher Scientific, UK), 2.5µl Better Buffer (Web Scientific, UK) and 0.5µl Big Dye Terminator v3.1 (Applied Biosystems, UK). All sequencing reactions were carried out following the manufacturers' recommendations (Applied Biosystems, UK). Sequences were visualised using an Applied Biosystems 3130 Genetic Analyzer.

The programs Sequencher 3.1.2. (Gene Codes Corporation, USA) and Bioedit (Hall 1999) were used to align and edit all sequences and to determine microsatellite haplotypes for each lineage.

Microsatellite fragment analysis[5]

L. rubellus individuals were collected from the Clydach study site, CL3 and Rudry study site, RU3 (Table 4.1) over the course of 2006-2007. These sites were located 77km apart and therefore could be considered allopatric populations. CL3 represented a rough pastureland site and RU3, an area of common land consisting of rough grassland. Earthworms were collected by digging and hand-sorting, and each individual was placed into a labelled polythene bag containing native soil.

Following collection, all earthworms were washed with distilled water and depurated over a period of two days. During this time the earthworms were kept inside petri-dishes containing two sheets of filter paper dampened with distilled water to prevent desiccation. The filter paper was changed after one day to allow the removal of excreted soil. After two days, earthworms were removed from the petri-dishes and rapidly frozen by immersion in a 277ml polystyrene beaker of liquid nitrogen. Once frozen, individual worms were placed into labelled plastic zip-loc bags and stored in a -80°C freezer. In order to prepare the worms for down stream processing, the frozen tissue was homogenised and ground into a fine powder using a mortar and pestle and liquid nitrogen to keep the tissue frozen. To prevent cross-contamination, all mortar and pestles were washed and sterilised using an auto-clave between samples. Again, the homogenised samples were stored in a -80°C freezer.

DNA was extracted from the frozen homogenised tissue samples using a DNeasy tissue extraction kit (QIAGEN, UK), following the manufacturers protocol for animal tissue extractions. Each extraction was assessed by gel electrophoresis, running 5µl of each sample and 1µl of loading buffer, using a 0.8% agarose gel stained with ethidium bromide. A 100 base pair DNA ladder was run as a control. All gels were visualised using a Gel-Doc-It Imaging System (UVP, UK) and the software, VisionWorks LS (UVP, UK). Any samples that failed to yield an observable band of DNA were re-extracted.

[5] See Fig 3.2 for workflow diagram of microsatellite fragment analysis

Mitochondrial sequence analysis of the cytochrome oxidase I region was performed to determine the mtDNA lineage of each individual (G. Juma pers. comm.).

The microsatellite primers applied in the study were derived from three different sources, namely the *L. rubellus* markers developed by Harper et al. (2006), primers developed for *Lumbricus terrestris* (Velavan et al. 2007) and microsatellite repeat regions identified within EST sequences (www.earthworms.org, Stürzenbaum et al. 2003). For the EST sequences, primers were designed using the program PRIMER 3 (Rozen and Skaletsky 2000).

In order to observe the amplification efficiency of all primer pairs, initial PCR reactions were performed in 10µl reaction volumes using 0.5µl of extracted *L. rubellus* lineage A or B DNA and 0.1µM of each primer added to a PCR mixture containing 1x PCR Buffer (Invitrogen, UK), 3mM Mg2+ (Invitrogen, UK), 0.2mM dNTP mix (dATP, dCTP, dGTP, dTTP) (Invitrogen, UK), PCR grade water (Fisher Scientific, UK) and 1u Taq polymerase (Invitrogen, UK). For the initial PCR reactions all samples were initially denatured for 5 minutes at 94ºC and then proceeded through 35 cycles containing a denaturation step at 94ºC for 1 minute, an annealing step at 50ºC for 1 minute and an extension step at 72ºC for 1 minute. After 35 cycles the samples underwent a final extension period of 72ºC for 5 min. All reactions were performed using a TC-412 thermal cycler (Techne). The success of the amplifications was determined through gel electrophoresis using 10µl of each sample and 1µl of loading buffer, and a 0.8% agarose gel stained with ethidium bromide. A 100 base pair DNA ladder was run as a control. All gels were visualised using a Gel-Doc-It Imaging System (UVP, UK) and the software, VisionWorks LS (UVP, UK).

Primer pairs that yielded a high degree of non-target amplification products were optimised by running further reactions at higher annealing temperatures. The annealing temperatures were increased in increments of two degrees for each subsequent optimising reaction until the non-target amplification was minimised. If non-target amplification could not be minimised the primer pair was discarded. Primer pairs that failed to successfully amplify were optimised by running further reactions with the temperature of the extension steps reduced. Extension temperatures were reduced by increments of two degrees for each subsequent optimising reaction until microsatellite loci were successfully amplified. If microsatellite loci could not be amplified the primer pair was discarded. If primer pairs from the EST microsatellite loci failed to amplify successfully, further primer pairs were developed from the flanking sequence and trialled again.

Eleven microsatellite loci were amplified in four multiplex reactions (Table 3.4). The results for three of the microsatellite loci (LTM026, 08286 and C2-1) were excluded from the final analysis due to a lack of successful amplifications. In preparation for each multiplex reaction, a 10x primer mix was prepared containing 2µl of each multiplexed primer. Multiplex PCR amplifications were

then performed using 0.5µl of extracted *L. rubellus* DNA in 10µl reaction volumes containing 1x Multiplex PCR Mastermix (QIAGEN), 1x primer mix and RNAse free water. All samples were initially denatured for 5 minutes at 94ºC and then proceeded through 35 cycles containing a denaturation step at 94ºC for 1 minute, an annealing step at a primer-specific temperature (see Table 3.4) for 1 minute and an extension step at a primer-specific temperature for 1 minute. After 35 cycles the samples underwent a final extension period at a primer-specific temperature for 5 min. All reactions were performed using a TC-412 thermal cycler (Techne).

Following amplification, all samples were diluted 1/10 with RNAse free water and prepared for fragment analysis with 0.5µl of each diluted sample placed into a 10µl loading cocktail including 9.25µl Hi-Di Formamide and 0.25µl ROX 500 Size Standard (Applied Biosystems). Fragment runs were then performed using an ABI 3130 Genetic Analyser. All fragment analysis results were visualised and scored using Genemapper 4.0 (ABI). Multiplex reactions that failed to produce scorable results across all loci, were reamplified and fragment analysed for a second time. Markers that failed to produce scorable results were reamplified as singleplex reactions. These reactions were performed in the same manner as the initial PCR reactions but using fluorescently labelled forward primers and GoTaq Hot Start Polymerase (Promega). Following amplification, 0.2µl of each reaction was multiplexed together with two other reactions featuring different fluorescent labels and placed into a 10.1µl loading cocktail including 9.25µl Hi-Di Formamide and 0.25µl ROX 500 Size Standard (Applied Biosystems). Fragment runs were again performed using an ABI 3130 Genetic Analyser.

Statistical analysis

Tests to identify incidences of gametic linkage disequilibrium and deviations from Hardy-Weinberg equilibrium frequencies were performed by Markov chain methods using Genepop-on-the-Web (http://genepop.curtin.edu.au/index.html, Raymond and Rousset, 1995) applying default settings. Three different statistics of genetic diversity were determined for each population, namely allelic richness and observed and expected heterozygosities. Population values of allelic richness were determined using FSTAT and observed and expected heterozygosities using Genepop-on-the-Web. The inbreeding coefficient, Fis was also determined for each population using FSTAT. Pairwise Fst values and probability values were calculated using GenAlEx. Benjamini-Hochberg correction was subsequently performed on the probability values to control for false discovery rate (Thissen et al. 2002). Mean interlineage and intralineage pairwise Fst values were calculated for each microsatellite locus and compared in a paired t-test using Minitab 16 (Minitab Inc).

An analysis of molecular variance (AMOVA) was conducted to determine whether the largest proportion of molecular variance was distributed between the different lineages, populations or individuals.

Multivariate factorial coordinate analysis (FCA) was performed using GENETIX 4.05.2 (Belkhir et al. 2004). This analysis enabled the visualisation of differences in allele frequencies between the individuals of different lineages and sites.

Bayesian clustering analysis was performed using the program STRUCTURE 2.3.3 (Pritchard et al. 2000). To ascertain the most likely number of clusters (K) within each data-set, initial runs were performed over K-values of 1-10 using three independent repeats. These initial runs were performed using an admixture model with a burn-in period of 10,000 repetitions followed by 100,000 MCMC repetitions. The most suitable value of K was then determined using the method proposed by Evanno et al. (2005) which assesses the change in likelihood value among values (delta K) as opposed to the maximum value of k as originally proposed. Following the determination of K, full runs were performed with a burn-in period of 100,000 repetitions followed by 1,000,000 MCMC repetitions.

Bayesian clustering analysis was also performed using the program NEWHYBRIDS. This program enables the identification of hybrid and back-cross individuals. No prior identification of the populations included in the study was required for the analysis. The analysis was run with a burn-in period of 5000 repetitions followed by 45,000 MCMC repetitions.

Table 3.1 Microsatellite multiplex reactions with repeat motifs, primer sequences, annealing temperatures and fluorescent dyes of individual microsatellites

Multiplex	Locus	Repeat motif	Primer sequence (5'-3')	Annealing temperature	Label
1	LTM163	TGC	F:GCCGGAGCGTTAGGAGCGATAG R:TTAGTGGTGAGTCGGGCGTATCC	52	FAM
1	9572	GAAT	F:ATGAAAATCAGGCGAATTGC R:GGGAAACGATTATGAGTCAAA	52	HEX
1	11297	TA	F:TATTCCGTTCATGAGTTTGAA R:CTCGTGTGTTGGCTTACGG	52	TAM
2	C3-2	CA	F:TCGAGCGCAGTTAGTTACGA R:GATCATGTGACTAACCAGATTA	55	FAM
2	G3-4	GA	F:ATTGGCGGACTTCAACAAAC R:CGGGTGTGCATTGTTTACAG	55	HEX
2	G2-1	GA	F:TCGAACTTCGAGTCTCTTCAA R:CGGTGTCGCTGATAGGTTTAG	55	NED
3	8286	GATT	F:TGAACCTGCTAATGCAACACA R:CACTGCCCTCACTTGGTACA	53	FAM
3	8138	GT	F:ATTGAAGCGCAGGATATTCG R:GGCGCTTTAGAAATGTGGAG	53	HEX
4	A3-2	CA	F:AAACTGTCGAAATGCACTCC R:CGGTTTGTCATCCCAATC	50	FAM
4	C2-1	CA	F:ATGATGAGTTTCGCCTGTTG R:AAGCTGGGGGTGATTCTTCT	50	HEX
4	LTM026	GT	F:GTGCCTCTGTCTAATGTCTGCTCGTGTGTA R:GCCGCTCTTTATACGCTCGTCGC	50	TAM

Table 3.2 Microsatellite libraries created for cloning

Library	Lineage	Locus
1	A	LTM163, 08138, 11297
2	A	LTM163, 08138, 11297
3	B	09572, C3-2, G3-4, G2-1
4	B	09572, C3-2, G3-4, G2-1

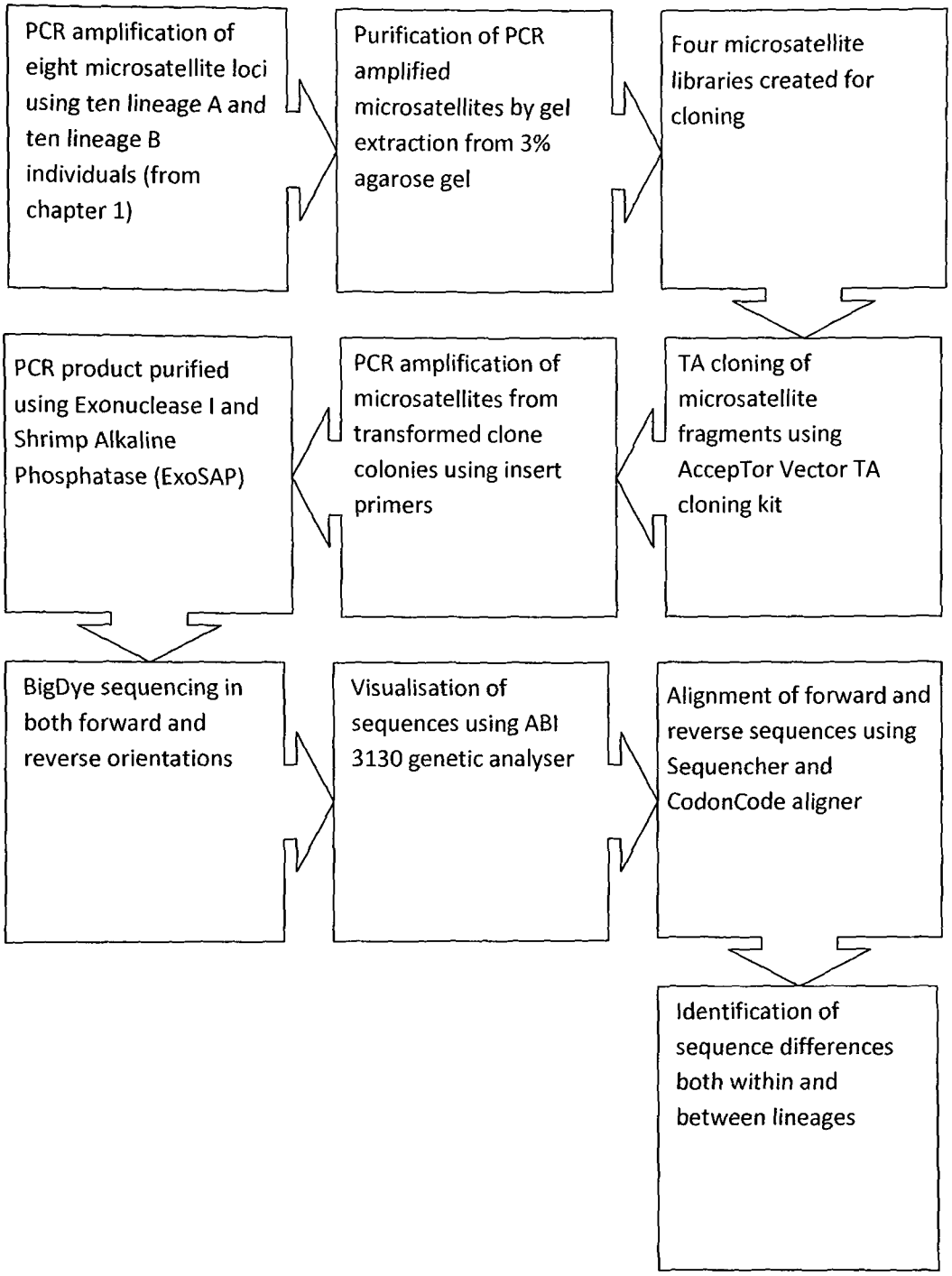


Figure 3.1 Work-flow diagram of microsatellite sequencing methodology

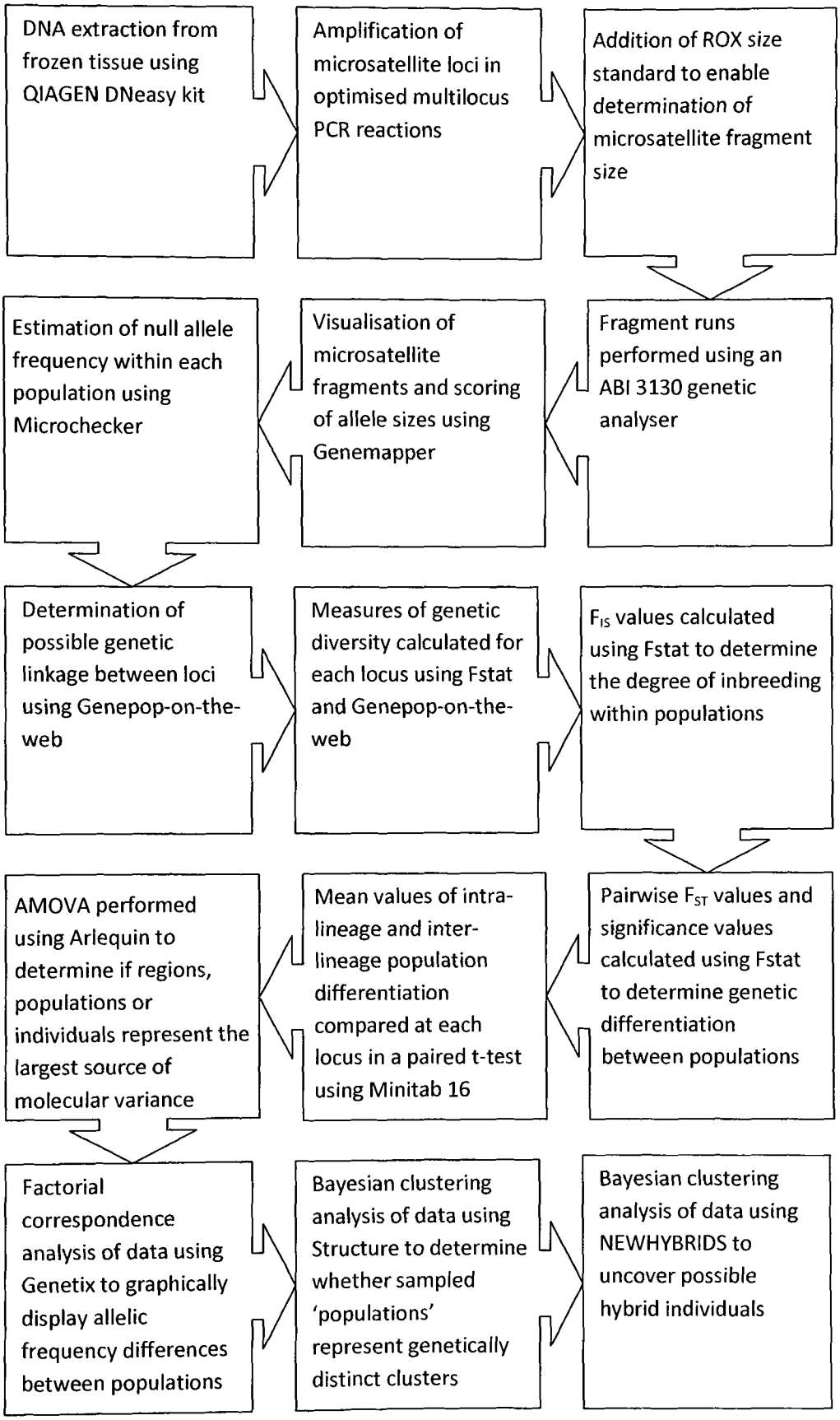


Figure 3.2 Work-flow diagram of microsatellite fragment analysis methodology

Results

Microsatellite fragment analysis

The microsatellite fragment analysis of 51 lineage 'a' and 'b' *L. rubellus* individuals from two different populations for eight loci revealed 48 alleles for lineage 'A' and 38 alleles for lineage 'B',. Within lineage 'A', the Clydach population displayed 39 alleles and the Rudry population displayed 27 alleles. Within lineage 'B', the Clydach population displayed 21 alleles and the Rudry population displayed 28 alleles. Several of the loci displayed monomorphism (only a single allele present) although this was largely confined to single populations. However, within lineage 'B' the locus G2-1 was found to be monomorphic for both populations. The allele size-ranges of three of the microsatellite loci were also found to be mutually exclusive between the two lineages (Table 3.3). The mean allelic richness value of 5.78 uncovered for lineage 'A' individuals was higher than the mean value for lineage 'B' (4.29). Within 'A', the higher value was found at Clydach (r=4.43. Rudry: r=3.33). Conversely, within 'B' the higher value was found at Rudry (r=3.32: Clydach: r=2.09).

Linkage analysis and Hardy-Weinberg equilibrium

Linkage analysis revealed only one significant association between loci (P=<0.05). This was within the Clydach lineage 'B' samples and was between the loci C3-2 and 08138. Within lineage 'A' some loci were found to display heterozygote deficiencies (Table 3.3). This was most pronounced at Clydach where six loci displayed lower heterozygosity than expected under Hardy-Weinberg equilibrium. The observed heterozygote frequencies for lineage 'B' differed less from expected values than lineage 'A'. However, with the exception of Clydach lineage 'A', all site Fis values were non-significant (Table 3.4).

Population differentiation

Pairwise Fst values indicated significant differentiation among all four populations analysed (Table 3.5). Importantly, the pairwise Fst values between the lineages were found to substantially exceed pairwise values between populations of the same lineage. An analysis of molecular variance based upon these Fst values ascribed the greatest proportion of variance to that existing between the two species (Table 3.6). However this was not statistically supported.

The mean pairwise Fst values of individual microsatellite loci ranged between 0.03 and 0.46 within the lineages and between 0.21 and 0.64 across the lineages. A paired t-test of the mean pairwise Fst values within and across the lineages for each microsatellite locus indicated a significant degree of differentiation between these values (P=0.03).

FCA analysis clearly differentiated lineage A and B individuals with no evidence of intermediary, hybrid genotypes (Fig 3.3). The two populations of lineage A were also more clearly differentiated than those of lineage B.

A determination of suitable K-values using the approach detailed by Evanno et al. (2005), supported a K-value of 2 for the total data-set (Fig 3.4). These two clusters were found to clearly differentiate individuals of the two *L. rubellus* mitochondrial lineages across both the Clydach and Rudry populations. No individuals were assigned to a different lineage to that indicated by their mtDNA genotype. This indicates that although the two lineages occur sympatrically at these sites they are strongly differentiated from one another with no evidence of male-mediated introgression.

Within lineage A, the individuals of Clydach and Rudry grouped into separate clusters at K=3. However a small degree of admixture was observed within the Clydach individuals with one individual found to display a higher level of assignment to the lineage A Rudry cluster (0.75). Within lineage 'B' the individuals of Clydach and Rudry were grouped into separate clusters at K=5 (Fig 3.5).

A Bayesian analysis using NewHybrids uncovered two clusters corresponding to lineages A and B. All individuals were designated as purebred genotypes of these two lineages (Fig 3.6). No evidence of hybrid or back-cross genotypes were uncovered during the analysis.

These results contrast with the findings of Andre et al. (2010) who uncovered several possible hybrid individuals. These individuals clustered with neither lineage during FCA analysis and instead were positioned in an intermediary position. No such individuals were uncovered in the present study.

Microsatellite flanking sequence variation

In total, 19 unique sequence haplotypes were found for the microsatellite flanking regions (MFRs) (Tables 3.8 and 3.9). Of these sequences, eight were obtained for lineage A and eleven for lineage B. Haplotypes from both lineage A and lineage B were obtained for all loci. Within lineage A, multiple MFR haplotypes were obtained for one of the microsatellite loci (C3-2) and for lineage B, multiple MFR haplotypes were obtained for three loci (08138, 09572, 11297). As expected, a high degree of sequence conservation was observed within primer-binding areas with the identification of only two substitutions across all loci. Both of these comprised single base-pair deletions and were observed within lineage B haplotypes of the markers 11297 and 09572.

Evidence of transitional, transversional and indel mutations were all found when comparing lineage A and B MFR haplotypes. The number of mutations between haplotypes of different

lineages was found to exceed those between haplotypes of the same lineage at all loci were such comparisons were possible, i.e. multiple haplotypes within lineages. The highest number of mutations found within a lineage was two, whereas the highest number of mutations between lineages was nine (Table 3.7). Five of the loci displayed indel mutations between lineages. These ranged in size from 1bp to regions of over 30bp. The immediate proximity of some of these mutations to the repeat regions may indicate that some of them may actually represent heavily corrupted repeat regions. These sections were therefore included as part of the repeat region sequences. Other mutations appear to represent nascent repeat regions as they have resulted in the loss/gain of small numbers of mononucleotide sequences. Two such examples of this are the indel mutations uncovered between lineages A and B for the locus 08138. Only two indel mutations were identified from intraspecific comparisons of MFR haplotypes for the loci 09572 and 11297. Both of these mutations represented sequence differences of 1bp.

Microsatellite repeat region variation

Observation of the microsatellite repeat regions revealed much evidence of imperfections within these sequences (Table 3.9). A large number of these imperfections appeared to result from substitutional mutations within repeat units. However it was not possible to determine whether some imperfections within dinucleotide regions resulted from substitutional mutations or alternatively were the result of multiple indel mutations that had corrupted the repeat motif. More clear evidence of insertion mutations were observed in some loci that featured large sections of non-repetitive DNA located within the repeat region. Some other loci displayed evidence of dinucleotide and mononucleotide compound repeats within the repeat regions.

Comparisons between the repeat units of the two lineages revealed that the majority of sequence imperfections were not shared between the lineages. This was particularly true of the larger insertional mutations. Large insertional mutations were observed within the repeat-regions of both lineages at the locus 11297, but differed markedly in their sequences suggesting that they represented independent mutational events. Mutations that were found to exist in both lineages were largely representative of single base-pair substitutions and therefore may have arisen independently.

Table 3.3 Microsatellite loci statistics (a= number of alleles, size range, r= allelic richness, He= expected heterozygosity, Ho= observed heterozygosity, Fis and HW= Hardy-Weinberg deviation P)

	a	size range (bp)	r	He	Ho	Fis	HW (P)
L. rubellus "A"							
8138	7	82-127	7.00	0.70	0.38	0.47	<0.01
9572	4	131-250	3.68	0.32	0.32	0.03	0.29
11297	4	205-244	3.82	0.58	0.21	0.64	<0.01
A3-2	4	150-217	4.00	0.44	0.06	0.87	<0.001
C3-2	6	158-259	5.66	0.59	0.33	0.45	<0.01
G2-1	14	156-250	13.19	0.90	0.72	0.20	<0.01
G3-4	3	98-105	2.96	0.25	0.27	0.08	0.36
LTM163	6	190-224	5.94	0.75	0.29	0.62	<0.001
Average	6		5.78	0.57	0.32	0.42	<0.01
L. rubellus "B"							
8138	10	143-257	8.75	0.81	0.82	-0.01	0.69
9572	5	126-260	4.60	0.55	0.28	0.51	<0.001
11297	4	224-243	3.85	0.42	0.15	0.64	<0.001
A3-2	4	172-189	3.66	0.54	0.46	0.15	0.33
C3-2	4	176-189	3.62	0.61	0.54	0.12	<0.01
G2-1	1	174	1.00	0.00	0.00		-
G3-4	6	186-208	5.49	0.63	0.66	-0.04	0.92
LTM163	4	127-188	3.34	0.22	0.23	-0.07	1.00
Average	4.75		4.29	0.47	0.39	0.19	<0.01

Table 3.4 Population microsatellite statistics (N= number of individuals, r=mean allelic richness, He= mean expected heterozygosity, Ho= mean observed heterozygosity, Fis, Hardy Weinberg deviation P)

Population	*N*	*r*	He	Ho	Fis	HW (P)
CL3 "A"	10	3.33	0.35	0.31	0.37	<0.001
RU3 "A"	10	4.19	0.56	0.41	0.06	0.54
CL3 "B"	13	2.59	0.45	0.43	0.00	0.87
RU3 "B"	18	3.32	0.37	0.37	-0.02	0.56

Table 3.5. Pairwise Fst values between populations

	CL3A	RU3A	CL3B	RU3B
CL3A	-			
RU3A	0.28**	-		
CL3B	0.41**	0.54**	-	
RU3B	0.41**	0.51**	0.23**	-

**P<0.01

Table 3.6. Analysis of molecular variance (AMOVA)

Source of variation	d.f.	Sum of squares	Variance components	%	P
Among lineages	1	26.24	0.51	57.45	0.33
Among populations within lineages	2	2.54	0.04	4.12	>0.001
Among individuals within populations	98	33.68	0.34	38.43	0
Total	101	62.45	0.89		

Table 3.7. Number of mutations observed at individual microsatellite loci (subst: substitutional mutations, indel: insertion/deletion mutations)

Locus	within lineage A			within lineage B			between lineages		
	subst	indel	Total	subst	indel	total	subst	indel	total
11297	-	-	-	0	1	1	5	4	9
8138	-	-	-	1	0	1	3	2	5
9572	-	-	-	1	1	0	1	2	3
C3-2	2	0	2	-	-	-	3	0	3
G2-1	-	-	-	-	-	-	4	3	7
G3-4	-	-	-	-	-	-	1	0	1
LTM163	-	-	-	-	-	-	4	3	7

Table 3.8 Microsatellite flanking regions (substitutional mutations are highlighted in red, indel mutations are highlighted in blue, location of repeat regions are highlighted in grey, boxed regions indicate primer binding site)

Locus	Flanking sequence
11297a	TTCAAACTCATGAACGGAATAATTTTAATTTTTGTCATGTTTTGTGTTTATGCAA
11297b1	TTCAAACTCATGAACGGAATAATTTTAATTTTTGTCATGTTTTGTGTTTATACAA
11297b2	TTCAAACTCATGAACGGAATAATTTTAATTTTTGTCATGTTTTGTGTTTATACAA
11297a	ACACTTTGTGATATACGGT-CATGTTTATAAATTGTGATTTAATGCTGAAACTTA
11297b1	ACACTTTGTGAAATACGGT-CATGTTTATAAATTGTGATTTAACGCTGAAACGTA
11297b2	ACACTTTGTGAAATACGGT-CATGTTTATAAATTGTGATTTAACGCTGAAACGTA
11297a	CTATGAACGGAATTTAGTATAGCCTACCGAGTAAATAGGTCTAATA~GG~TTTTG
11297b1	CTATGAACGGAATTTAATATA~~~~~CCGAGTAAATAGGTCT~ATACGGTTTTTG
11297b2	CTATGAACGGAATTTAATATA~~~~~CCGAGTAAATAGGTCT~ATACGGTTTTTG
11297a	GCCGTAAGCCAACACACGAG
11297b1	GCCGTAAGCCAACACACGAG
11297b2	GCCGTAAG~CAACACACGAG
08138a	ATTGAAGCGCAGGATATTCGAGTGTTTT~CCATATCAATGCC~TAAAGTAATGTA
08138b1	ATTGAAGCGCAGGATATTCGAGAGTTTTTCCATATCAATGCCCTAAAGTAATATA
08138b2	ATTGAAGCGCAGGATATTCGAGAGTTTTTCCATATCAATGCCCTAAAGTAATGTA
08138a	TTTCCAACACAAAGTGTGCTCTCGACATTGTCGAATGTGCAGTTGATTTGTC-CG
08138b1	CTTCCAACACAAAGTGTGCTCTCGACATTGTCGAATGTGCCGTTGATTTGTC-CG
08138b2	CTTCCAACACAAAGTGTGCTCTCGACATTGTCGAATGTGCCGTTGATTTGTC-CG
08138a	TAATCGTAATCTCCACATTTCTAAAGCGCC
08138b1	TAATCGTAATCTCCACATTTCTAAAGCGCC
08138b2	TAATCGTAATCTCCACATTTCTAAAGCGCC
09572a	GGGAAACGATTATGAGTCAAAGCGATCTTCTAATGTTAGTAGTAGTCCCACTATG
09572b1	GGGAAACGATTATGAGTCAAAGCGATCTTCTAATGTTAGTAGTAGTCCCACTATG
09572b2	GGGAAACGATTATGAGTCAAAGCGATCTTCTAATGTTAGTAGTAGTCCCACTATG
09572b3	GGGAAACGATTATGAGT~AAAGCGATCTTCTAATGTTAGTAGTAGTCCCACTATG
09572a	CTT-AGGCCTATGAGTTAAGCGATCTTCTAATGTTAGTAGTAGTCCCACTATGCT
09572b1	CTT-AGGC~~~~~~~TTAAT~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
09572b2	CTT-AGGC~~~~~~~TTAAT~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
09572b3	CTT-AGGC~~~~~~~TTAAT~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
09572a	AAAAACGCAAGTAAATTAGTCTGGCGAAGATTCACACGCTACTTAATGTTCATGA
09572b1	~~~~~~GCAAGTAAATTAGTCTGGCGAAGATTCACACGTTACTTAATGTTCATGA
09572b2	~~~~~~GCAAGTAAATTAGTCTGGCGAAGATTCACACGCTACTTAATGTTCATGA
09572b3	~~~~~~GCAAGTAAATTAGTCTGGCGAAGATTCACACGTTACTTAATGTTCATGA
09572a	AAATCAGGCGAATTGC
09572b1	AAATCAGGCGAATTGC
09572b2	AAATCAGGCGAATTGC
09572b3	AAATCAGGCGAATTGC

Table 3.8 (continued) Microsatellite flanking regions (substitutional mutations are highlighted in red, indel mutations are highlighted in blue, location of repeat regions are highlighted in grey, boxed regions indicate primer binding site)

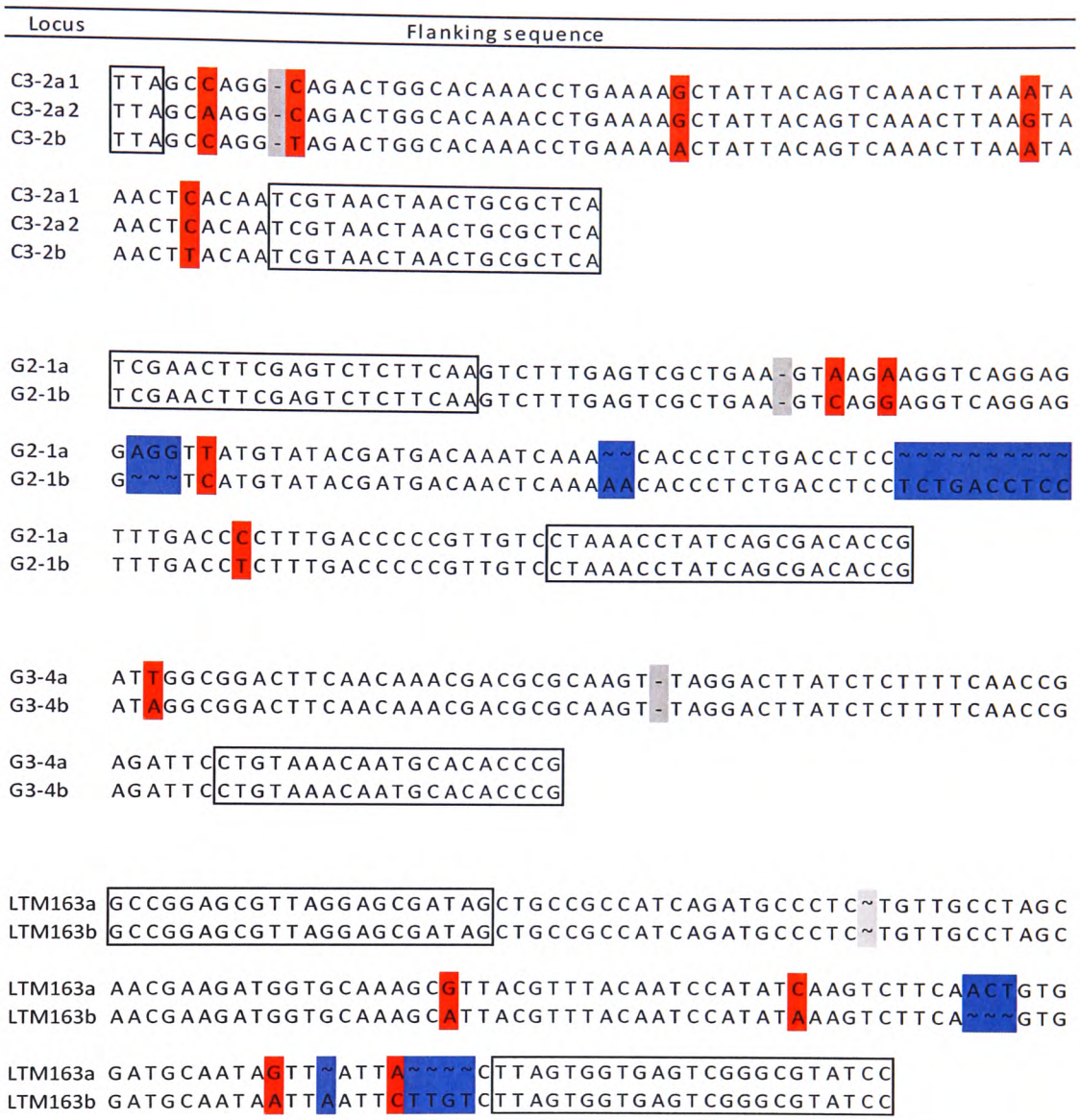


Table 3.9 Microsatellite repeat regions

Locus	Repeat sequence
C3-2a1/a2	(CA)10 AACAGA (CA)5 GACTGG (CA)2 AAGA (CA)5 GA (CA)6 GACAGG (CA)4 TG (CA)2 TA (CA)4 CG
C3-2b	(CA)2 TA (CA)4
08138a	(GT)12 TT (GT)4 TT (GT)9
08138b1	GTGA (GT)16
08138b2	(GT)18
09572a	AATAAT (GAAT)5
11297a1	TAGGCC (TA)3 A (TA)3 TTTTAAATGTATGGGCTAAT (TA)2 CA (TA)2 ACATATCTTACA
11297a2	TAGGCC (TA)3 A (TA)3 TTTTAAATG (TA)2 GGCTAAT (TA)2 CA (TA)2 ACATATCTTACA
11297b1	(TA)6 (CA)3 (TA)4 ATGTTCTTT (TA)3 G
11297b2	(TA)7 (CA)3 (TA)4 ATGTTCCTT (TA)3 G
11297b3	(TA)7 (CA)3 (TA)4 ATGTTCTTT (TA)3 G
11297b4	(TA)7 (CA)3 (TA)4 ATGTTCTTT (TA)3 G
G2-1a	(GA)4 AA (GA)4 Check end
G2-1b1	(GA)2 AA (GA)6
G2-1b2	(GA)2 AA (GA)6
G3-4a1	(GA)3 GC (GA)6 GC (GA)3
G3-4a2	(GA)3 GC (GA)5 GC (GA)3
G3-4b1	(GA)4 GC (GA)5 CAGC (GA)4 GC (GA)5 GC (GA)9 GC (GA)12 GC (GA)4 GGGC (GA)5 GC (GA)3
G3-4b2	(GA)4 GT (GA)5 CAGC (GA)4 GC (GA)5 GC (GA)9 GC (GA)12 GC (GA)4 GGGC (GA)5 GC (GA)3
LTM163a	(TGC)3 TGT (TGC)8
LTM163b	TCC (TGC)5
08138a	(GT)12 TT (GT)4 TT (GT)9
08138b1	GT GA (GT)16
08138b2	(GT)18

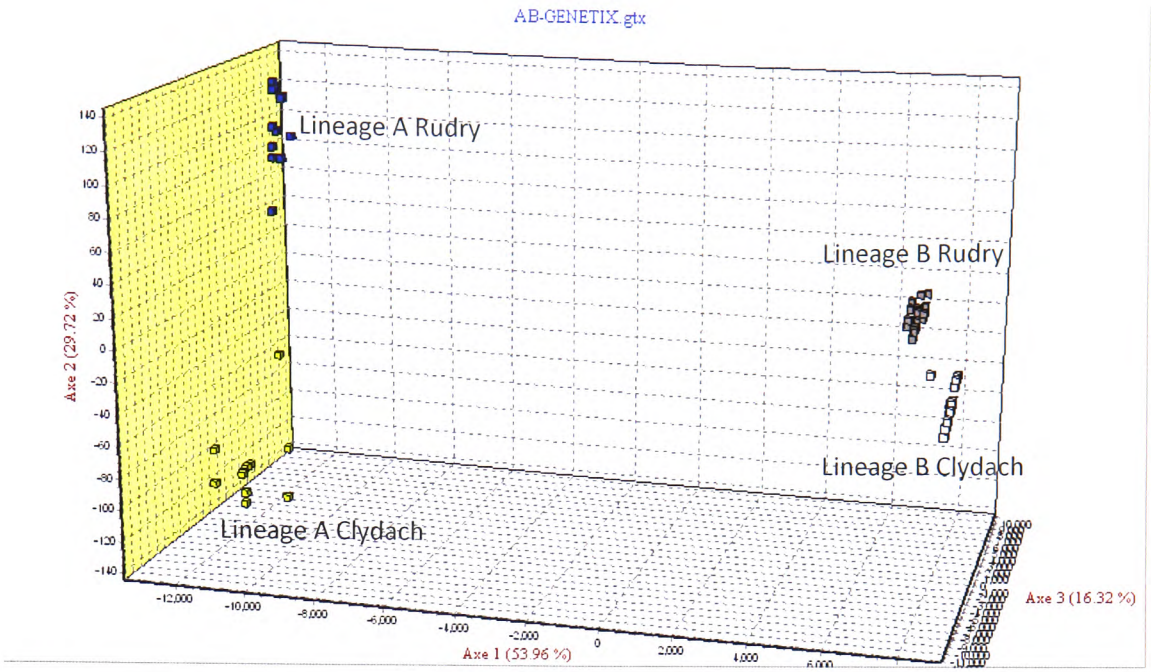


Figure 3.3 Three-dimensional FCA plot of Clydach and Rudry populations of lineage A and B *L. rubellus*. This plot graphically demonstrates the allelic differences between individuals in a multidimensional space

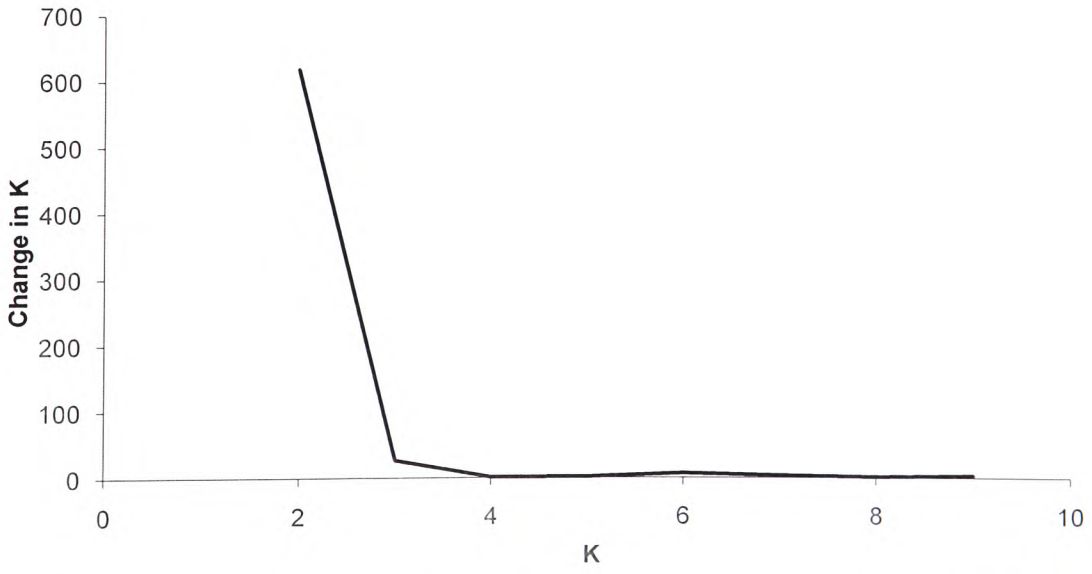


Figure 3.4 Determination of the most likely value of K (number of genetically distinct clusters) for the present data-set following the method of Evanno et al. (2005). In this method the optimal value of K displays the highest second order rate of change in log-probability value (designated Change in K on the graph).

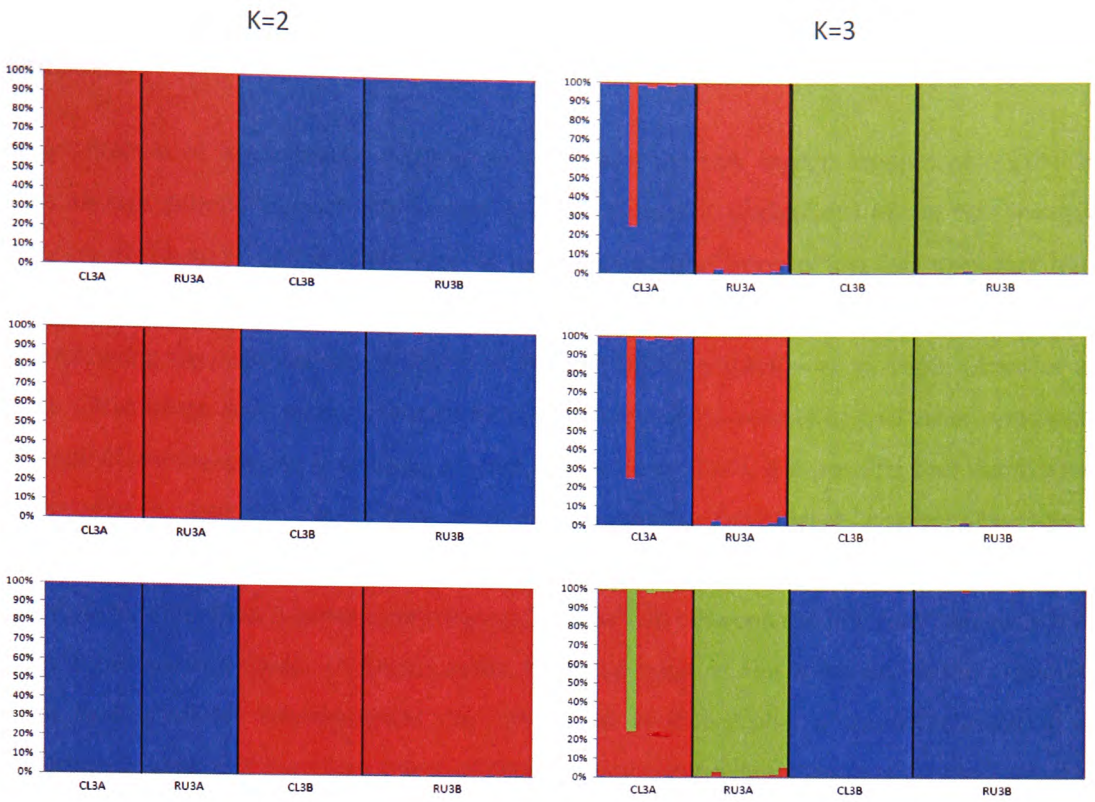


Figure 3.5 Results of Bayesian clustering analysis using STRUCTURE. This analysis determines whether sampled 'populations' correspond to distinct genetic clusters (K=number of clusters in the analysis)

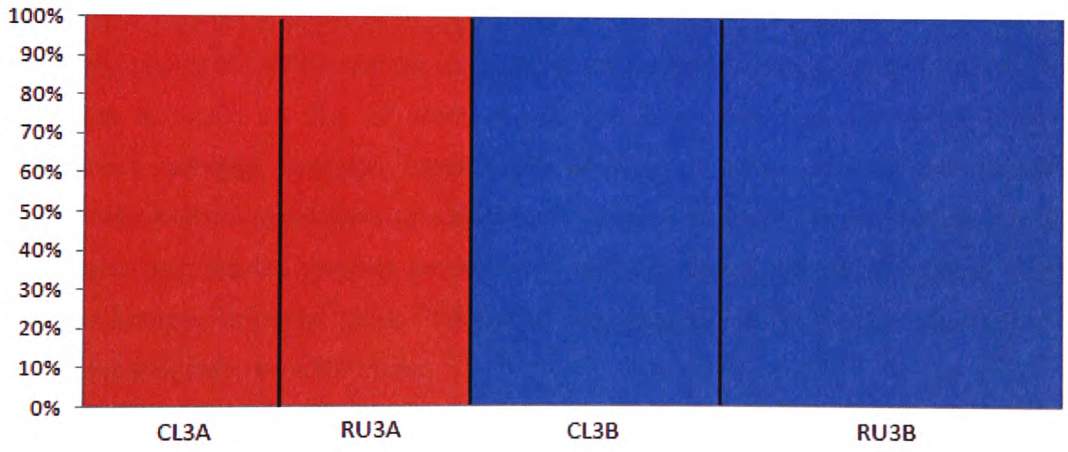


Figure 3.6 Results of Bayesian clustering analysis using NewHybrids (This analysis identifies pure-bred, hybrid and back-cross individuals)

Discussion

The application of microsatellite analysis to investigate the two cryptic lineages of *L. rubellus* indicated high genetic differentiation exists between sympatric populations of the two lineages. Firstly an observation of the allele sizes of loci revealed that three of the loci displayed non-overlapping size ranges. Such differences may be indicative of a general divergence in overall allele sizes following the reproductive isolation of the two lineages (Molbo et al. 2003, Elmer et al. 2007). Also given that many interspecific comparisons of microsatellite loci have uncovered evidence of size homoplasy (Estoup et al. 1995, Fisher et al. 2000), it is possible that the five loci that overlap in their size ranges may result from a genetic convergence in allele size. This is supported by the observation that many of the alleles that were common within one lineage were absent within the other. A genetic convergence in allele size between the different lineages would be expected under models of microsatellite evolution such as the stepwise mutation model. Under such a model, microsatellites can either gain or lose repeat units. After a period of divergence in allele size therefore, microsatellite loci may eventually converge to an allele size that was present prior to the divergence occurring.

A pairwise analysis of allele frequencies indicated that differentiation between sympatric populations of the two lineages as indicated by Fst values, far exceeded that between geographically separate populations of a single lineage. This difference between inter-lineage and intra-lineage population differentiation was found to be statistically supported by a paired t-test. This finding is similar to that of other studies that have uncovered far higher levels of differentiation between sympatric populations of closely related species than between geographically isolated populations of a single species (Li et al. 2009). An analysis of molecular variance also ascribed the greatest proportion of variance within the results to that existing among the lineages. However this was not statistically supported which could be due to the low number of demes featured in the study.

Bayesian clustering analysis of the allele frequencies also revealed a strong differentiation of the two lineages. Cluster analysis using the program STRUCTURE strongly supported the division of the sampled individuals into two clusters corresponding to the two lineages. A further analysis using the program NEWHYBRIDS also strongly supported a division of the data into two clusters. Furthermore the program was unable to detect any evidence for the existence of hybrid or back-cross individuals between the two lineages. This contrasts with the study of Andre et al. (2010), who also applied nuclear genetic markers to study the two lineages of *L. rubellus*. In the study, a multi-locus amplified fragment length analysis uncovered a small number of possibly hybrid individuals within two sites where both lineages were found to exist sympatrically and two other sites which were found to be composed predominately of a single lineage. However the relatively

small number of possible hybrids uncovered in the study, did however suggest that the hybridisation of the two lineages may be a relatively rare event, or one forced by extreme demographic or environmental perturbation.

Sequencing of microsatellite alleles from the two lineages uncovered many substitutions between lineages within the flanking regions. For the loci where comparisons of intra- and inter-lineage sequences substitutions were possible, a higher level of sequence divergence was identifiable between the alleles of different lineages than for those of the same lineage. The identification of mutations within the microsatellite flanking regions of the two lineages of *L. rubellus* supports the view that the lineages are deeply divergent and may have separated over a long evolutionary time period. Past studies that have applied microsatellite flanking sequences to assist in the resolution of phylogenies have discovered that mutations within the MFRs of conserved microsatellites are relatively rare, resulting in a rate of evolution which is comparable to that observed between some mitochondrial DNA sequences. Mutations within MFRs are therefore more likely to be identified during comparisons of sequences from different taxa than they are during comparisons of sequences from conspecific individuals. A high degree of genetic divergence between the two lineages of *L. rubellus* is also indicated by the discovery of a greater proportion of insertion/deletion mutations between lineages than within lineages. This again is in agreement with other studies of microsatellite flanking regions that have found insertion and deletion mutations to be less frequent than substitution mutations, with the largest indel mutations often found during comparisons over deep taxonomic levels (Rico et al. 1996, Zardoya et al. 1996).

The inter-lineage comparison of sequence differences within the repeat regions of some of the microsatellite loci also indicated the existence of a substantial degree of divergence between the lineages. Many of the microsatellite repeats were found to include imperfections including the non-repeat sequences and compound repeats. As in the flanking regions, some of these mutations were shared between haplotypes within a single lineage suggesting an origin subsequent to the divergence of these two lineages.

Overall, the discovery of a clear genetic differentiation between populations of the two *L. rubellus* lineages and the identification of microsatellite sequence differences between them indicates reproductive isolation between the two lineages. Given the highly conserved nature of microsatellite flanking sequences that has been indicated by existing studies, the discovery of mutations within these sequences between lineages indicates that they may have been separated for a long evolutionary time-period. These findings support the inferences of a high degree of genetic differentiation between the lineages obtained by mtDNA analysis (King et al. 2008, Andre et al. 2010).

There are several ways in which reproductive isolation may currently exist between the two lineages. Firstly mating studies of some congeneric earthworm species have demonstrated the existence of post-zygotic isolation between species in the form of hybrid inviability, with cross-species matings resulting in the formation of inviable cocoons (Domínguez et al. 2005). Hybrid inviability has also been demonstrated in breeding experiments featuring *Allolobophora chlorotica* (Lowe and Butt 2008), a species that that has been found to contain highly divergent cryptic genetic lineages (King et al. 2008). Breeding experiments between some of the lineages was found to result in the formation of a high degree of inviable cocoons. The study of Lowe and Butt (2008) also indicated that some lineages may be reproductively isolated due to differences in their optimal environmental conditions for growth, with the individuals of one lineage found to demonstrate a higher degree of fitness within wet soils.

As King et al. (2008) remark, genetically distinct cryptic lineages within a species are more usually found to exist in allopatry. Several studies of cryptic variation within earthworm species have found evidence for the existence of genetically differentiated allopatric populations (Chang et al. 2008, Novo et al. 2010b). It is therefore possible that the two lineages within *L. rubellus* may have originated in allopatry and have subsequently become sympatric in their distributions. One possibility is that the two lineages originated from two different glacial refugia and mixed during recolonisation following the end of the Pleistocene (King et al. 2008). This theory has been considered unlikely given that most of the genetic lineages that are known to have developed from peri-glacial refugia now occur in allopatry (Taberlet et al. 1998). Another possible explanation is that one of the lineages may have survived the last glaciations within a cryptic refugium close to its present geographical location (Stewart and Lister 2001). Mixing of the two lineages may have then occurred following the end of the Pleistocene.

Another possibility is that the two lineages may have diverged in sympatry. This may have occurred due to a divergence in the ecological niche of the two lineages. Gut content analysis and feeding trials have indicated that earthworm species of the same broad ecological niche (e.g. epigeic, endogeic) can often display further differences in their feeding ecology (Curry and Schmidt 2007). This includes a preference for different types of leaf litter (Satchell 1967), different particle sizes of soil (Piearce 1978) and possibly different species of fungus or bacteria (Curry and Schmidt 2007). It is therefore possible that the two lineages may display differences in their feeding ecology. Alternatively the lineages may display differences in their preference for certain soil conditions such as moisture like the differences observed between lineages of *Allolobophora chlorotica* (Lowe and Butt 2008).

It has been proposed that many sympatric species that display differences in their ecological niche may have evolved sympatrically through a process of divergent selection (Bolnick 2004, Elmer et

al. 2010). Studies indicate that reproductive isolation may subsequently act to prevent the formation of sub-optimally adapted hybrid individuals (Nosil et al. 2003). It is therefore possible that the two lineages of *L. rubellus* may have diverged from one another by such a process resulting in their reproductive isolation.

Chapter 4. Population genetics analysis of *Lumbricus rubellus* populations along a gradient of nickel contamination using microsatellite markers

Although many earthworm species have been found to represent ideal sentinel species for assessing the effects of soil contamination, some studies have indicated that earthworm populations at contaminated sites may be more resistant towards certain contaminants than individuals from clean sites. Earthworms collected from sites contaminated with arsenic (Langdon et al. 1999, Langdon et al. 2001), cadmium (Voua Otomo and Reinecke 2010) and copper (Arnold et al. 2007) have all demonstrated a higher degree of tolerance towards their respective contaminants than have worms from non-contaminated control sites during laboratory exposure trials. Such resistance is indicated by the exposed earthworms displaying a lower degree of mortality (Langdon et al. 1999, Reinecke et al. 1999, Langdon et al. 2001), better overall condition (Langdon et al. 1999, Langdon et al. 2001) and less evidence of cytotoxic and genotoxic effects at low concentrations of contaminants (Voua Otomo and Reinecke 2010). Indications that such resistance may have a genetic basis rather than, or in addition to, being a physiological adaptation or maternal effect (see Posthuma and Van Straalen 1993) have been suggested by studies that have demonstrated resistance in laboratory-bred F1 and F2 offspring of earthworms from contaminated sites (Langdon et al. 2009). Laboratory-based studies featuring earthworms collected from non-contaminated sites have also demonstrated that an increased resistance may develop following only a few generations of breeding within contaminated soil (Spurgeon et al. 2000). It is not apparent however whether the increased resistance that has been observed is a consequence of being genetically adapted or alternatively because they are more phenotypically plastic than non-resistant individuals (Langdon et al. 2009).

It has been proposed that the process of adaptation to environmental stress could result in the development of major demographic changes within populations (Posthuma and Van Straalen 1993, Belfiore and Anderson 2001, Van Straalen and Timmermans 2002). By analysis of the population genetics of exposed and non-exposed populations it may therefore be possible to infer the presence of genetically adapted populations at contaminated sites. Some advantages of this approach to the inference of genetic adaptation are that it can incorporate the signal of past demographic events (e.g. genetic bottlenecks) and that it offers a greater degree of resolution by focusing upon an entire population rather than the reactions of several individuals (Belfiore and Anderson 2001). However, population genetic diversity and structure may also be affected by factors other than adaptation to contaminants. For example, several studies that have observed patterns of neutral genetic differentiation between contaminated and reference sites have found evidence of genetic structuring that is more likely the result of geographical factors than contamination related stress (Whitehead et al. 2003, McMillan et al. 2006). It is therefore important that any studies of the genetics of exposed populations are designed in such a way as to

allow the effects of adaptation to be disentangled from other possible demographic effects. There are two possible ways by which geographical differences can be accounted for in studies of contamination, namely the study of multiple contaminated sites or of sites located along a gradient of contamination (Posthuma and Van Straalen 1993). Such an approach makes it easier to correlate demographic effects with the presence of specific contaminants within a site.

In this study we aimed to infer whether populations of *Lumbricus rubellus* inhabiting a nickel contaminated environment are genetically resistant. In order to enable the inference of genetically adapted populations, individuals were sampled along a gradient of nickel contamination. Patterns of genetic diversity and differentiation were determined for the populations along the gradient, by genotyping microsatellite loci. Microsatellites are regions of repetitive DNA that are believed to be largely selectively neutral (although see- Li et al. 2002) and which are widely applied in the field of population genetics to enable inferences of genetic diversity and population differentiation (Beaumont and Bruford 1999, Chistiakov et al. 2006). Microsatellite markers have already been applied within studies aiming to uncover the population effects of anthropogenic contaminants upon plants (Muller et al. 2007), fish (Silbiger et al. 2001, Whitehead et al. 2003, Bourret et al. 2008, Knapen et al. 2009) and mammals (Berckmoes et al. 2005). Microsatellite markers have also been successfully applied within a small number of general phylogenetic studies of earthworm species (although these do not feature populations from contaminated sites) (Velavan et al. 2009, Novo et al. 2010a, Somers et al. 2011).

Several studies have demonstrated that microsatellite analysis can uncover evidence of population differentiation resulting from selection. The differentiation discovered by such studies has often been found to be in agreement with data from other molecular markers or studies of phenotypic differentiation. Some of the most compelling evidence for adaptation to environmental stress obtained using microsatellite data has been uncovered within several studies of the freshwater fish *Poecilia mexicana* (Plath et al. 2007, Tobler et al. 2008, Plath et al. 2010). These studies analysed genetic differentiation between populations of fish from drainage systems featuring populations exposed to both subterranean and sulfidic environments. This has revealed a high degree of differentiation between populations of different environments and has also identified these separate populations as discrete groupings during cluster analysis. The findings of such microsatellite analyses are concordant with the findings of phenotypic studies in suggesting that populations from different types of environment are adapted to their respective habitats, resulting in a low degree of gene flow between these populations. An incidence of flooding within the cave-system also supported the presence of selection within different environments (Plath et al. 2010). Following the flooding an increased level of gene flow was inferred between populations within the same environment. However no changes in gene flow were observed between

populations from different environments, indicating the presence of strong selection for adapted phenotypes. Another microsatellite study which has indicated the presence of selection is that of Hendry et al. (2002) which focused upon several populations of the freshwater fish, *Gasterosteus aculeatus* from different environments within a lake system. This revealed a high degree of genetic differentiation to exist between populations from the inlet and the lake itself. These populations were also found to display a high degree of morphological differentiation indicating the possible selection of these populations to their respective habitats.

The sensitivity of microsatellite data therefore makes it suitable for analyzing the demographic effects of a contamination gradient of nickel upon populations of the earthworm *Lumbricus rubellus*. The environment under investigation has been contaminated by the emissions from a nickel smelting works for over a century (Grimsrud and Peto 2006). Prevailing winds have forced emissions in a north-easterly direction resulting in the formation of a contamination gradient of nickel along the Swansea valley (Morgan, pers. Comm.).

This project therefore assessed the patterns of genetic variation within and among populations of the earthworm, *L. rubellus* and whether selective pressures resulting from the pollution have driven the adaptation of earthworms resulting in the development of nickel-tolerant populations. Demographic change among populations was inferred using microsatellites to enable the identification of genetic diversity and differentiation between individual earthworms and sampling sites. Analysis of these differences enabled patterns of gene flow between sites and populations to be inferred and any alteration in dispersal patterns, potentially due to soil contamination, to be identified. In order to enable the disentanglement of demographic effects resulting from contamination and geographical factors, *L. rubellus* populations were also sampled at two non-contaminated control sites, an approach that has been recommended for population studies of contamination (Posthuma and Van Straalen 1993). At the reference site of Rudry, earthworms were sampled within a series of horizontally stratified soils that differ in pH. Studies of earthworm abundance have indicated that soil pH may have a major impact upon the demographics of earthworm populations (Chan and Mead 2003, Hirth et al. 2009).

Two general scenarios and predictions were tested during the investigation, although it is recognised that these are not mutually exclusive:

Hypotheses

1. Samples of *L. rubellus* on nickel-contaminated land are sustained by the local recruitment of genetically adapted individuals. These individuals display an increased degree of nickel-tolerance and are therefore able to successfully survive and reproduce in their (contaminated) environment.

2. Samples of *L. rubellus* on contaminated land represent 'demographic sinks', comprising recent immigrants, that have dispersed from less contaminated land. These individuals may have a reduced level of fitness through living in contaminated conditions but may also be able to survive in their contaminated surroundings through a process of phenotypic plasticity.

Predictions

A. Following from 1) above, samples taken closest to the contaminated site should be the most genetically differentiated and possess the lowest within-sample genetic diversity, either as a result of local adaptation, founder effect or demographic bottlenecks (Van Straalen and Timmermans 2002).

B. Following from 2) above, samples taken closest to the contamination source should be the most admixed and therefore possess the highest within-population genetic diversity and the lowest average genetic differentiation.

A related prediction to B), but unrelated to 2) is that nickel pollution increases genetic diversity via mutation. It has been proposed that some genotoxic compounds may raise gene diversity by inducing heritable genetic mutations (Bickham et al. 2000). In this case we would predict that genetic diversity within samples living in contaminated land should be high but that private alleles (unsampled elsewhere) would also be expected to be most frequent in these samples.

Materials and Methods

Sample collection

L. rubellus individuals were largely collected from all three of the study sites, namely Clydach, Rudry and Dinas Powys over the course of 2006-2007 (Figs 4.1-4.4, See Appendix A for geographical locations). At the nickel contaminated site of Clydach earthworms were collected along a pollution gradient located downwind of the pollution point source (INCO Nickel Smelter, Clydach) with distances ranging between approximately 200m to 5.8km away (Fig 4.5). The geographic locations of the sampling stations were also selected to allow possible barriers to gene flow along the gradient to be taken into account during the analysis. These included the River Tawe and the main roads. A further round of sampling was also conducted during April 2008, during which an extra thirty earthworms were collected from a twelfth sampling station located at Clydach. The inclusion of an extra sampling station at Clydach was deemed necessary because only nine *L. rubellus* individuals were found at Clydach station 5 due to a low overall abundance of earthworms at that particular sampling station. The sites at Clydach represented a range of different habitat types that included urban sites, woodland and agricultural land (see Appendix B for site details). In addition to Clydach, earthworms were also collected from ten sampling stations at the horizontally-stratified control site of Rudry (Fig 4.6) and nine sampling stations at the second control site of Dinas Powys (Fig 4.7). At Rudry, the sample stations were selected so as to represent communities from a range of different soil types (Table 4.1). Most of the sample sites of Rudry and Dinas Powys were located within areas of woodland, non-improved grassland or pastureland (see Appendix B for site details).

At each sampling station earthworms were collected by digging and hand-sorting, and each individual was placed into a labelled polythene bag containing native soil. The location of each site was recorded using an eTrex GPS receiver (Garmin, UK). The dry soil weight of nickel at each site was determined by acid extraction of metals from the soil samples (performed by Dr Gabriela Juma). This revealed the sites of CL01 and CL05 to be nickel contaminated in accordance with soil guideline values established by the Environment Agency (Environment Agency 2009). Both sites displayed nickel dry weight values of over 1000 mg/kg (See Appendix B). From here on in nickel contaminated sites will be indicated with a superscript 'N'.

DNA extraction

DNA was extracted from the frozen homogenised tissue samples using a DNeasy tissue extraction kit (QIAGEN, UK), following the manufacturers protocol for animal tissue extractions. Each extraction was assessed by gel electrophoresis, running 5µl of each sample and 1µl of loading buffer, using a 0.8% agarose gel stained with ethidium bromide. A 100 base pair DNA ladder was

run as a control. All gels were visualised using a Gel-Doc-It Imaging System (UVP, UK) and the software, VisionWorks LS (UVP, UK). Any samples that failed to yield an observable band of DNA were re-extracted.

Mitochondrial sequence analysis of the cytochrome oxidase I region was performed to determine the mtDNA lineage of each individual (G. Juma pers. comm.).

Microsatellite genotyping[6]

Eleven microsatellite loci were amplified in four multiplex reactions (Table 4.2). The development and optimisation of these microsatellite loci is described in chapter 3. The results for two of the microsatellite loci (LTM026 and C2-1) were excluded from the final analysis due to a lack of successful amplifications. In preparation for each multiplex reaction, a 10x primer mix was prepared containing 2µl of each multiplexed primer. Multiplex PCR amplifications were then performed using 0.5µl of extracted *L. rubellus* DNA in 10µl reaction volumes containing 1x Multiplex PCR Mastermix (QIAGEN), 1x primer mix and RNAse free water. All samples were initially denatured for 5 minutes at 94ºC and then proceeded through 35 cycles containing a denaturation step at 94ºC for 1 minute, an annealing step at a primer-specific temperature (see Table 3.4) for 1 minute and an extension step at a primer-specific temperature for 1 minute. After 35 cycles the samples underwent a final extension period at a primer-specific temperature for 5 min. All reactions were performed using a TC-412 thermal cycler (Techne).

Following amplification, all samples were diluted 1/10 with RNAse free water and prepared for fragment analysis with 0.5µl of each diluted sample placed into a 10µl loading cocktail including 9.25µl Hi-Di Formamide and 0.25µl ROX 500 Size Standard (Applied Biosystems). Fragment runs were then performed using an ABI 3130 Genetic Analyser. All fragment analysis results were visualised and scored using Genemapper 4.0 (ABI). Multiplex reactions that failed to produce scorable results across all loci, were reamplified and fragment analysed for a second time. Markers that failed to produce scorable results were reamplified as singleplex reactions. These reactions were performed in the same manner as the initial PCR reactions but using fluorescently labelled forward primers and GoTaq Hot Start Polymerase (Promega). Following amplification, 0.2µl of each reaction was multiplexed together with two other reactions featuring different fluorescent labels and placed into a 10.1µl loading cocktail including 9.25µl Hi-Di Formamide and 0.25µl ROX 500 Size Standard (Applied Biosystems). Fragment runs were again performed using an ABI 3130 Genetic Analyser.

[6] See Fig 4.8 for work-flow diagram of microsatellite fragment analysis

Statistical analysis

During data analysis, analyses were performed separately for two different data-sets. The first of these consisted of the allelic data from all nine microsatellite loci at all sample sites. The second data set consisted of allelic data from eight microsatellite loci at all sample sites to allow additional data from lineage A individuals sampled at sites CL03 and RU03 (see chapter 3).

Null allele frequencies at each loci were estimated using the program MICROCHECKER (Van Oosterhout et al. 2004). Tests to identify incidences of gametic linkage disequilibrium and deviations from Hardy-Weinberg equilibrium frequencies were performed by Markov chain methods using Genepop-on-the-Web (http://genepop.curtin.edu.au/index.html, Raymond and Rousset, 1995) applying default settings.

Three different statistics of genetic diversity were determined for each population, namely allelic richness and observed and expected heterozygosities. Population values of allelic richness were determined using FSTAT and observed and expected heterozygosities using Genepop-on-the-Web. Comparisons of allelic richness, observed and expected heterozygosity and Fis were also performed between all three of the regions in permutation tests using FSTAT (1000 permutations). The number of private alleles within each population at Clydach was determined using CONVERT (Glaubitz 2004) in order to observe the effects of nickel contamination upon genetic diversity.

The inbreeding coefficient, Fis was also determined for each population using FSTAT. Pairwise Fst values and probability values were calculated using GenAlEx. A Benjamini-Hochberg correction for multiple-comparisons was performed on the probability values to control for false discovery rate (Thissen et al. 2002). This particular correction was applied as it is less conservative than other applicable methods such as Bonferroni correction.

An analysis of molecular variance (AMOVA) was conducted to determine whether the largest proportion of molecular variance was distributed between different regions, populations or individuals. To supplement this AMOVA analysis, pairwise Fst values between sites from different regions were compared to those between sites from within single regions. An unpaired two-way t-test was conducted for the 9-loci data-set using Minitab with the pairwise Fst values arcsine square-root transformed to normalise the data sets. Results for the 8-loci data-set could not be normalised by transformation and so were analysed using a non-parametric Mann-Whitney test.

Pairwise Fst values were also tested for an isolation-by-distance effect within each of the three regions using the Isolde test implemented by Genepop-on-the-Web. The test was performed with Fst data converted to values of Fst/(1-Fst) and using the natural log of the geographical distances.

The Isolde test utilised a one-tailed Spearman Rank correlation test followed by a Mantel test of 1000 permutations. Pairwise Fst values were also tested for correlation with the pairwise differences between sites for soil nickel concentration within Clydach (Table 4.10) and soil calcium (as a measurement of pH) within Rudry (Table 4.11). Again these tests were performed using the Isolde tests implemented by Genepop-on-the-web. The relationship between the genetic differentiation of the sites and their level of nickel contamination (for Clydach) and soil pH (for Rudry) were further investigated in a second analysis. In this analysis the mean pairwise Fst value between each site and every other site in the region was calculated for both the Clydach and the Rudry samples. Correlations between these mean pairwise values with soil nickel for Clydach and soil pH levels (measured as soil calcium concentration) for Rudry were then tested using a regression analysis in Minitab.

Multivariate factorial coordinate analyses (FCA) of both the entire data-set and the individual regions were performed using GENETIX 4.05.2 (Belkhir et al. 2004). These analyses enabled the visualisation of differences in allele frequencies between the individuals of different populations and regions.

Bayesian clustering was performed using the program STRUCTURE 2.3.3 (Pritchard et al. 2000). Analyses performed included a cluster analysis of the whole data-set, the three individual regions, individual soil types for the Rudry sites and proximity to the smelter for the Clydach sites. To ascertain the most likely number of clusters (K) within each data-set, initial runs were performed over K-values of 1-10 using three independent repeats. These initial runs were performed using an admixture model with a burn-in period of 10,000 repetitions followed by 100,000 MCMC repetitions. The most suitable value of K was then determined using the method proposed by Evanno et al. (2005). Following the determination of K, second runs were performed with a burn-in period of 100,000 repetitions followed by 1,000,000 MCMC repetitions.

Table 4.1. Soil types of Rudry sites as indicated by reference to a geological map

Site	Soil type
RU01	Clay
RU02	Clay
RU03	Coal
RU05	Limestone
RU06	Limestone
RU08	Millstone grit

Table 4.2 Microsatellite multiplex reactions with repeat motifs, primer sequences, annealing temperatures and fluorescent dyes of individual microsatellites

Multiplex	Locus	Repeat motif	Primer sequence (5'-3')	Annealing temperature	Label
1	LTM163	TGC	F:GCCGGAGCGTTAGGAGCGATAG R:TTAGTGGTGAGTCGGGCGTATCC	52	FAM
1	9572	GAAT	F:ATGAAAATCAGGCGAATTGC R:GGGAAACGATTATGAGTCAAA	52	HEX
1	11297	TA	F:TATTCCGTTCATGAGTTTGAA R:CTCGTGTGTTGGCTTACGG	52	TAM
2	C3-2	CA	F:TCGAGCGCAGTTAGTTACGA R:GATCATGTGACTAACCAGATTA	55	FAM
2	G3-4	GA	F:ATTGGCGGACTTCAACAAAC R:CGGGTGTGCATTGTTTACAG	55	HEX
2	G2-1	GA	F:TCGAACTTCGAGTCTCTTCAA R:CGGTGTCGCTGATAGGTTTAG	55	NED
3	8286	GATT	F:TGAACCTGCTAATGCAACACA R:CACTGCCCTCACTTGGTACA	53	FAM
3	8138	GT	F:ATTGAAGCGCAGGATATTCG R:GGCGCTTTAGAAATGTGGAG	53	HEX
4	A3-2	CA	F:AAACTGTCGAAATGCACTCC R:CGGTTTGTCATCCCAATC	50	FAM
4	C2-1	CA	F:ATGATGAGTTTCGCCTGTTG R:AAGCTGGGGGTGATTCTTCT	50	HEX
4	LTM026	GT	F:GTGCCTCTGTCTAATGTCTGCTCGTGTGTA R:GCCGCTCTTTATACGCTCGTCGC	50	TAM

Figure 4.1 INCO nickel smelter near sites $CL1^N$ and $CL5^N$

Figure 4.2 Rudry Common viewed from site RU3

Figure 4.3 Dinas Powys (site DP2)

Figure 4.4 Dinas Powys (site DP2)

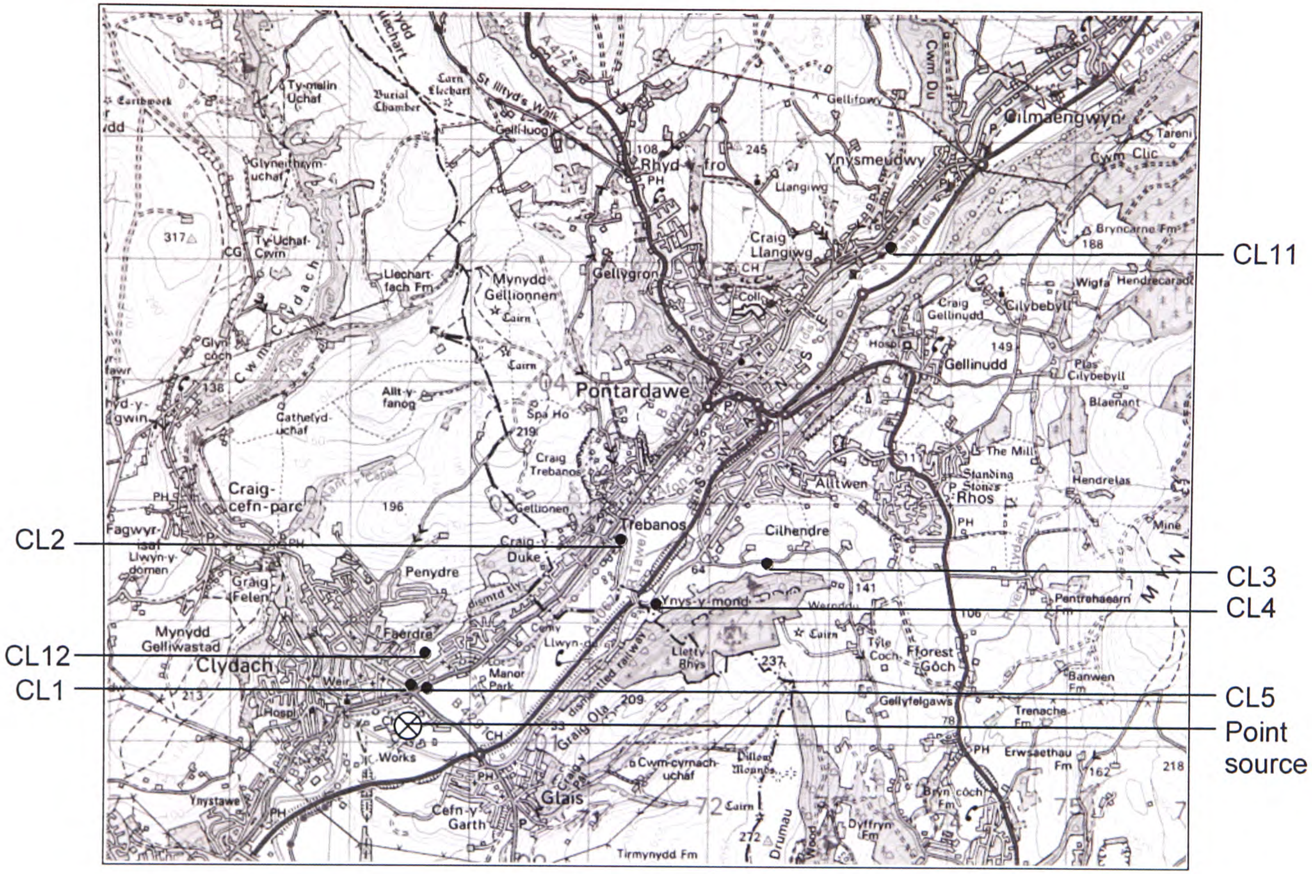


Figure 4.5 Location of sample sites at Clydach

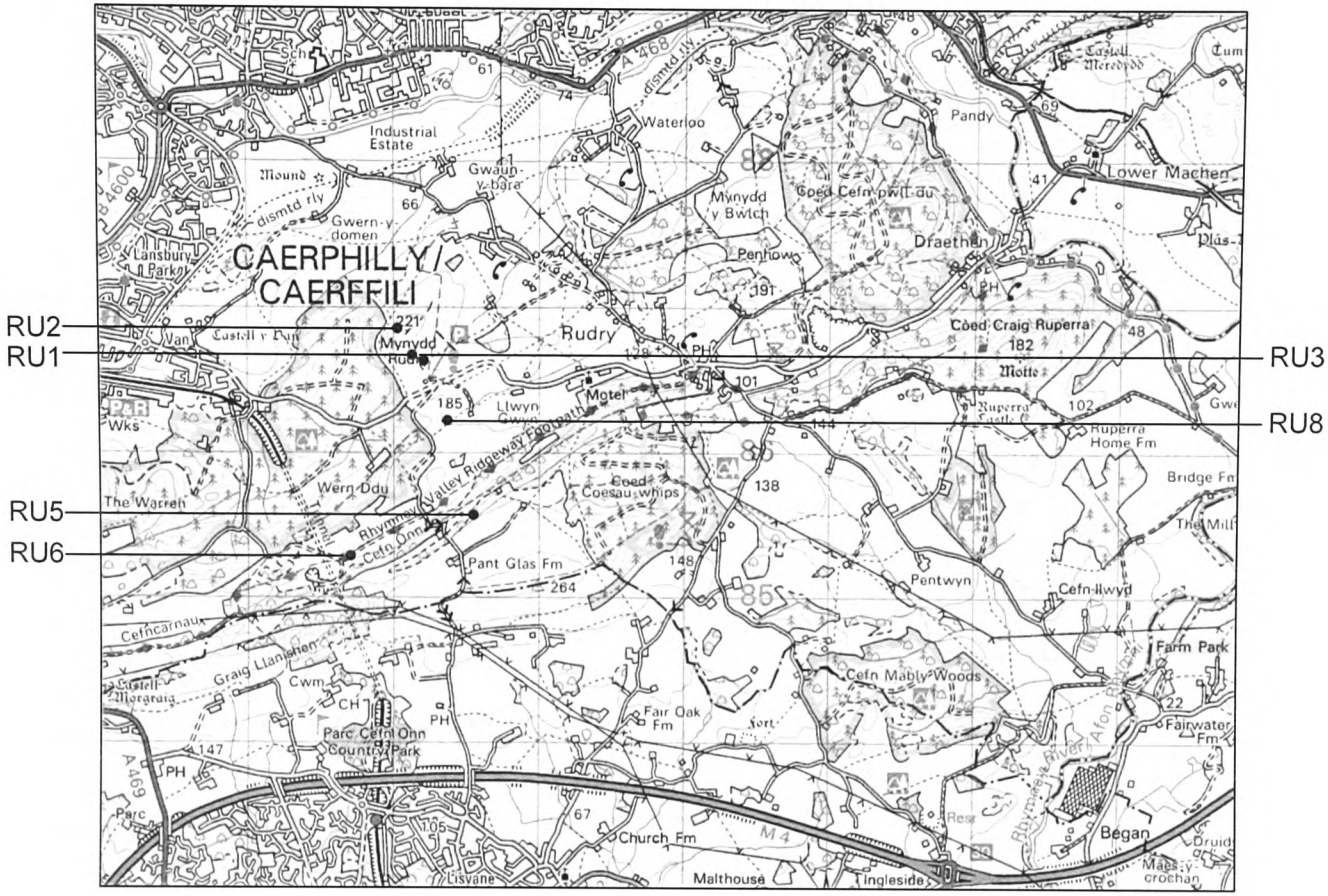


Figure 4.6 Location of sample sites at Rudry

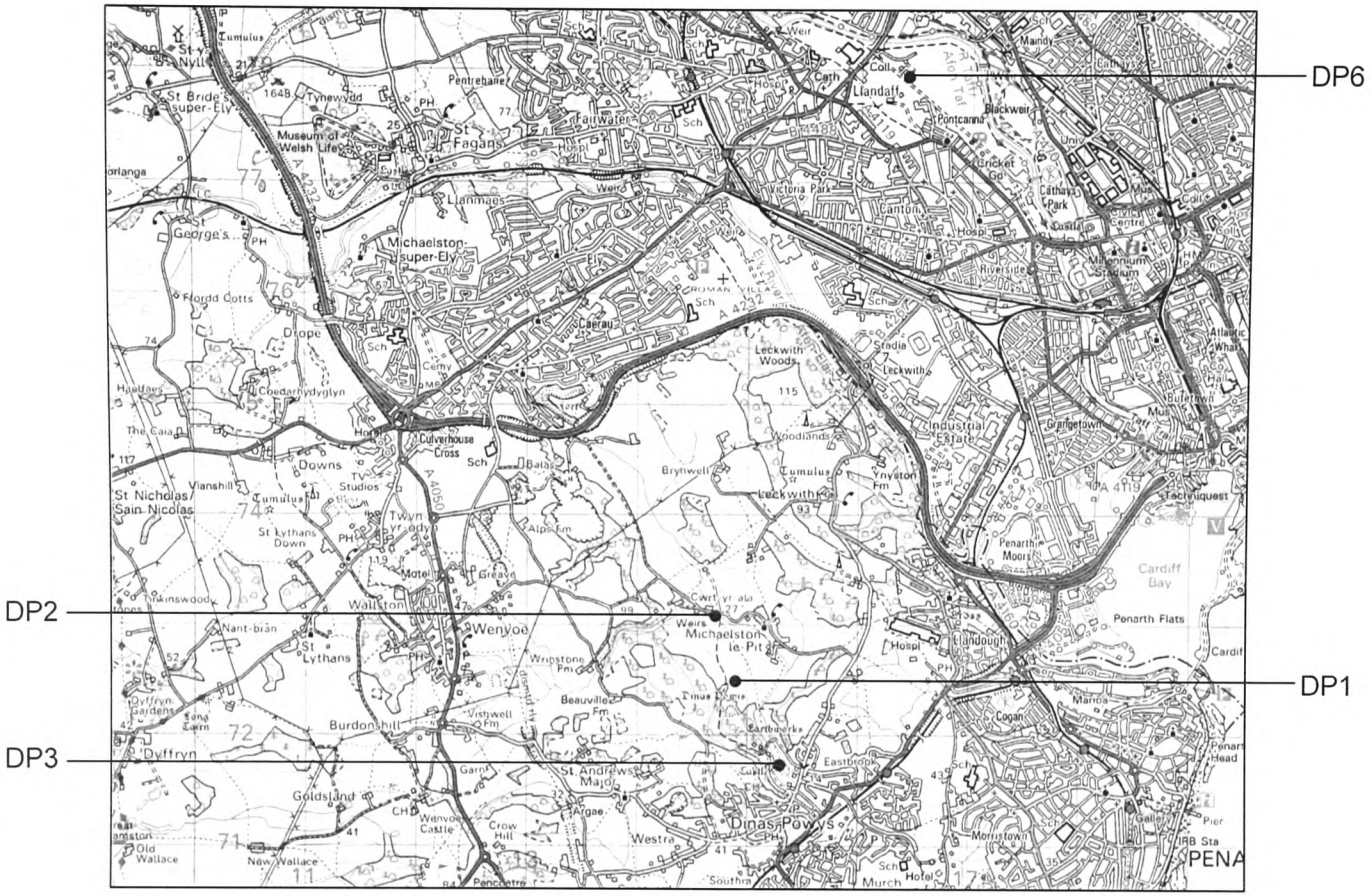


Figure 4.7 Location of sample sites at Dinas Powys

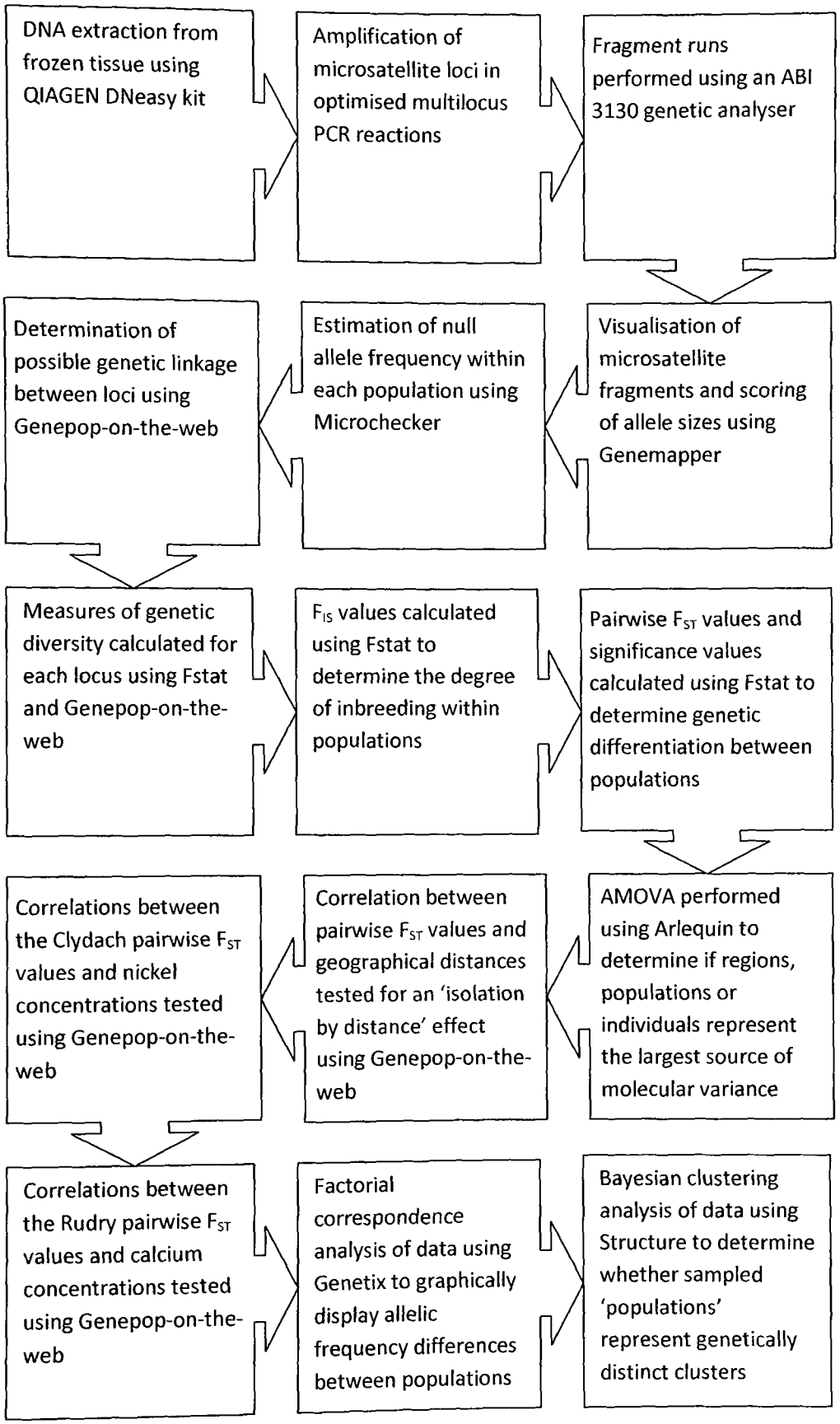


Figure 4.8 Work-flow diagram of microsatellite fragment analysis methodology

Results

Fragment analysis

Fragment analysis of *L. rubellus* populations from all three study regions over nine loci revealed 147 alleles in total. Analysis of the Clydach populations revealed the highest number of alleles for the three different regions with 99 alleles identified across six sample sites (87 individuals). Analysis of the Dinas Powys populations identified 98 alleles across four sample sites (79 individuals) and analysis of the Rudry populations identified 85 alleles across five samples sites (109 individuals). Levels of average allelic diversity ranged between 3.22 and 6.78, and allelic richness between 1.66 and 6.46 (Table 4.3). The lowest values of both average allelic diversity and richness were obtained for the site $CL05^{N}$ which was represented by only five individuals. One-tailed comparisons of allelic richness between populations of the three different regions using permutation tests revealed that the populations of Rudry displayed a significantly lower degree of allelic richness than those of both Clydach ($p=0.001$) and Dinas Powys ($p=0.002$). No significant difference was found between the levels of allelic richness for Clydach and Dinas Powys.

The analysis of *L. rubellus* populations over eight microsatellite loci to enable the addition of extra study sites for Clydach and Rudry revealed 139 alleles in total. As before, the highest number of these alleles was uncovered within the Clydach populations with 100 alleles identified across seven sample sites (97 individuals). Analysis of the Dinas Powys populations identified 82 alleles across four sample sites (79 individuals) and analysis of the Rudry populations identified 78 alleles across six sample sites (119 individuals; Table 4.4). Average allelic diversity ranged between 3.22 and 6.78, and allelic richness between 1.64 and 6.28 (Table 4.4). The lowest values of both average allelic diversity and richness were again obtained for the site $CL05^{N}$. One-tailed comparisons of allelic richness between populations of the three different regions revealed that the populations of Rudry displayed a significantly lower degree of allelic richness than Clydach ($p=0.001$). No significant difference was found between the levels of allelic richness for Clydach and Dinas Powys or Rudry and Dinas Powys.

Of the nine microsatellite loci analysed, the most variable was found to be the locus G2-1 with 39 different alleles. The least variable microsatellite locus was found to be the locus A3-2 which was monomorphic in two of the Rudry populations.

Within the Clydach sites the number of private alleles was found to range between 1 and 8. Two of the highest numbers of private alleles were found for sites close to the smelter ($CL01^{N}$ and CL12) (Table 4.5).

Linkage analysis and Hardy-Weinberg equilibrium

Within the 9-loci analysis, linkage analysis revealed 13 significant associations (P=<0.05) between loci within 8 sample sites. Within the 8-loci analysis, 10 significant associations were uncovered within 8 sample sites. However, as no consistent pattern could be observed in either the distribution of these associations or the loci affected, it was considered unlikely that they reflected evidence of linkage disequilibrium between loci or population processes such as admixture.

Hardy-Weinberg analysis of individual microsatellite loci revealed evidence of significant heterozygote deficiencies for seven of the microsatellite loci in the study across several sites. Large Fis values were found across all sites, ranging between 0.14 and 0.47 for the 9-loci dataset and 0.14 and 0.49 for the reduced 8-loci dataset. All of the sample Fis values were found to differ significantly from zero in the 9-loci dataset. However in the 8-loci data set, Fis values for the populations of $CL05^{N}$ and RU03 were not significant.

Analysis of heterozygote frequencies using MICROCHECKER indicated the possible existence of null alleles at loci found to exhibit high levels of homozygosity. However the adjustment of allele frequencies to include estimated null allele frequencies failed to prevent deviations from Hardy-Weinberg equilibrium. No further analysis was therefore performed using the transformed data-set.

Genetic differentiation

Pairwise genetic differences between the three different regions ranged between 0.03 and 0.29 for the 9-locus analysis (Table 4.6) and 0.03 and 0.35 for the 8-locus analysis (Table 4.7). Following Benjamini-Hochberg correction two pairwise values within the 8-locus data-set were found to be non-significant (P=>0.05) (DP01-RU02, DP01-RU06). Overall the pairwise genetic differentiation between sites from different regions appeared to exceed that between sites from the same region. Both an unpaired t-test of the 9-locus data-set and a Mann-Whitney test of the 8-locus data-set indicated these pairwise differences to be highly significant (P=<0.001). However some of the sites did appear to contradict this overall pattern of differentiation. Within Clydach the site of CL03 appeared more highly differentiated from some of the other sites within this region than it was from some of those of Dinas Powys. Similarly, within Dinas Powys, the site of DP03 was found to be more differentiated from the other sites within this region than it was from some of the Clydach sites.

AMOVA analysis indicated that the highest proportion of molecular variance existed at the individual level for both data-sets (9 locus: 73.3% P=<0.001, 8 locus: 73.5% P=<0.001) (Table 4.8).

However the AMOVA also revealed that 21.2% (P=<0.001) of the variance in the 9-locus analysis and 19.8% (P=<0.001) of the variance in the 8-locus analysis was distributed among the three different regions. This was far higher than the variance among sites within these groups (9 locus= 5.6% P=<0.001, 8 locus= 6.7% P=<0.001).

In Clydach pairwise Fst values ranged between 0.00 and 0.09 for the 9-locus analysis and between 0.01 and 0.16 for the 8-locus analysis. For the 9-locus analysis four of the pairwise values within Clydach were found to be non-significant following Benjamini-Hochberg correction (p=>0.05) (CL02-CL04, CL02-CL05^{N}, CL11-CL02, CL12-CL05^{N}). For the 8-locus analysis only one of the pairwise values was found to be non-significant (CL04-CL05^{N}). It is possible that these non-significant values may result from the low number of individuals genotyped for populations CL02 and CL05^{N}.

No clear patterns of genetic differentiation between the samples could be observed within either of the Clydach data-sets. Although the geographical distance between sampled sites at Clydach ranged between 60m and 5.6km, there was little evidence of a correlation between the pairwise genetic differentiation and geographical distance between samples (Fig 4.9). Pairwise distances between some of the most geographically proximate sites e.g. CL01^{N} and CL05^{N}, were found to be higher than those between some of the more geographically distant samples e.g. CL01^{N} and CL11. A lack of correlation between genetic differentiation and geographical distance was underlined by the Mantel tests of isolation by distance which indicated the correlation to be non-significant in both data-sets (9 locus P=0.30, 8 locus P=0.39).

The pairwise genetic differentiation values did not correlate with pairwise differences in nickel concentration between sites (Fig 4.10). Analysis of the correlation using a Mantel test indicated the relationship to be non-significant for both data-sets (9 locus P=0.88, 8 locus P=0.45). A second analysis of this correlation which compared mean pairwise Fst values with soil nickel concentrations did indicate a slightly higher degree of mean differentiation within two of the most heavily contaminated sites in the 9-locus analysis (CL01^{N} and CL05^{N}) compared with less contaminated sites. However the differentiation between these two sites and the less contaminated sites was not as pronounced within the 8-locus analysis, possibly due to the addition of data representing the highly differentiated CL03 site. The regression analysis of both data-sets did not reveal significance for either, although the p-value for the 9-locus data-set was not far from significance (9 locus P=0.06; 8 locus P=0.93).

Within Rudry pairwise Fst values ranged between 0.00 and 0.07 for the 9-locus analysis and between 0.00 and 0.06 for the 8-locus analysis. Following Benjamini-Hochberg correction only one pairwise value within the 9-locus data-set was found to be non-significant (p=>0.05) (RU01-RU02).

No clear patterns of genetic differentiation between sites could be observed within either of the Rudry data-sets. As at Clydach there was little evidence of a correlation between the pairwise genetic differentiation and geographical distance between sites (Fig 4.9) with Mantel tests of isolation by distance indicating the correlation to be non-significant in both data-sets (9 locus P=0.68, 8 locus P=0.77). It is worth mentioning however that the geographical scale of sampling across Rudry was smaller than across the other two regions, with pairwise geographical distances ranging between 140m and 1.6km.

The pairwise genetic differentiation of sites within Rudry also did not appear to be clearly correlated with differences in soil calcium content (Fig 4.11). A Mantel test indicated this correlation to be non-significant for both of the Rudry data-sets (9 locus P=0.57, 8 locus P=0.54). A second analysis of this correlation featuring the regression analysis of mean pairwise Fst values and soil calcium concentration was also non-significant for both data-sets (9 locus P=0.38, 8 locus P=0.35).

Within Dinas Powys Fst values ranged between 0.01 and 0.06 (using 9-locus). Following Benjamini-Hochberg correction only one pairwise value was found to be non-significant (P=>0.05) (DP01-DP02).

The pairwise genetic differentiation of sites within Dinas Powys did not correlate with geographical distance (Fig 4.9). As an indication of this the two most genetically differentiated sites of Dinas Powys were also the second closest in terms of their pairwise geographical distance (DP01-DP03). A Mantel test of isolation by distance was found to be non-significant (P=0.37).

Population structure

An FCA analysis of all sites clearly differentiated individuals from the three different regions (Fig 4.12). This differentiation did however appear to be counter-intuitive when geographic distances between the three regions were considered. The individuals of Clydach and Dinas Powys appeared to be less differentiated from one another than they were from Rudry. This was surprising when it is considered that these two regions are both located closer to Rudry than each other.

A determination of suitable K-values using the approach detailed by Evanno et al. (2005) indicating the most highly supported number of distinct genetic clusters within the data-set supported a K-value of 2 (Fig 4.13). This resulted in the grouping of all individuals into two distinct clusters (Fig 4.14), one cluster consisting of the Rudry individuals and the other consisting of Clydach and Dinas Powys individuals. At K=3 individuals were further differentiated into three clusters representative of the three different sampling regions.

Analysis of the Clydach sites revealed strong support for a K-value of 4 when analysing individuals from seven sites with nine microsatellite loci (Fig 4.13). However, no clusters were identified across any of the site, with a high degree of admixture within all sites (Fig 4.15). A separate analysis of the Clydach sites including CL03 was conducted based upon the allelic data from eight microsatellite loci. Here, strong support for a K-value of 5 (Fig 4.13) was found. At K=5, the individuals of CL03 formed a cluster that was differentiated from other sites (Fig 4.16). The clustering of CL03 individuals was also found to be discernable at K values as low as K=2. A separate analysis of the three sites in closest proximity to the Clydach smelter with nine microsatellite loci supported a K-value of 3 (Fig 4.13). This identified a cluster consisting of all individuals from the $CL01^{N}$ site (Fig 4.17). A separate analysis of sites further downwind of the smelter based upon eight microsatellite loci, was unable to produce a clearly defined value of K for the four sites analysed (Fig 4.13). However, the individuals of CL03 were found to form a distinct cluster at K=2 (Fig 4.18).

Analysis of the Rudry individuals revealed strong support for a K-value of 4 when analysing five sites with all nine microsatellite loci (Fig 4.13). At this particular K-value no clusters were clearly discernible (Fig 4.19). However, at lower values of K some evidence of clustering was apparent. At K=2 the individuals of RU08 appeared to form a distinct cluster that was differentiated from other Rudry sites. At K=3 both the individuals of RU05 and RU08 appeared to form distinct clusters. A separate analysis of the Rudry sites including RU03 was conducted based upon allelic data from eight microsatellite loci. No value of K was found to be strongly supported during initial analysis of the data (Fig 4.13). However runs at both K=2 and K=3 were both found to demonstrate the clustering of individuals at RU08 (Fig 4.20). Individual analyses grouping together sites of the four different soil types did not uncover any evidence of clustering (Figs 4.21-4.24).

Analysis of the Dinas Powys individuals was unable to uncover strong support for any of the K-values analysed (Fig 4.13). Observation of the results for K=2-10 revealed no evidence for clustering at any of the K-values that were considered (Fig 4.25).

Table 4.3 Microsatellite statistics with 9 loci (N= number of individuals, r=mean allelic richness, He= mean expected heterozygosity, Ho= mean observed heterozygosity, Fis, Hardy Weinberg deviation P)

Site	N	r	He	Ho	Fis	HW (P value)
CL05[N]	5	1.66	0.62	0.53	0.14	0.07
CL01[N]	20	4.36	0.59	0.47	0.20	<0.001
CL12	22	6.08	0.68	0.47	0.32	<0.001
CL02	10	4.16	0.64	0.43	0.32	<0.001
CL04	12	4.26	0.64	0.49	0.24	<0.001
CL11	18	6.13	0.68	0.44	0.36	<0.001
Clydach- all	87	4.44	0.64	0.47	0.26	<0.001
RU01	19	3.45	0.42	0.30	0.40	<0.001
RU02	22	4.32	0.61	0.45	0.28	<0.001
RU05	24	4.32	0.61	0.45	0.28	<0.001
RU06	21	4.32	0.61	0.45	0.28	<0.001
RU08	23	4.58	0.61	0.44	0.29	<0.001
Rudry- all	109	4.20	0.57	0.42	0.31	<0.001
DP01	22	5.42	0.64	0.39	0.47	<0.001
DP02	22	6.09	0.63	0.41	0.34	<0.001
DP03	19	6.46	0.69	0.53	0.30	<0.001
DP06	16	5.76	0.68	0.42	0.41	<0.001
Dinas- all	79	5.93	0.66	0.44	0.36	<0.001
All	550	4.78	0.62	0.44	0.31	<0.001

Table 4.4 Microsatellite statistics with 8 loci (N= number of individuals, r=mean allelic richness, He= mean expected heterozygosity, Ho= mean observed heterozygosity, Fis, Hardy Weinberg deviation P)

Site	N	r	He	Ho	Fis	HW (P value)
$CL05^N$	5	1.64	0.59	0.49	0.14	0.05
$CL01^N$	20	4.41	0.57	0.46	0.21	<0.001
CL12	22	5.96	0.66	0.45	0.33	<0.001
CL02	10	3.93	0.62	0.43	0.29	<0.001
CL04	12	4.08	0.62	0.47	0.25	<0.001
CL03	10	3.33	0.35	0.31	0.37	<0.001
CL11	18	6.28	0.67	0.43	0.36	<0.001
Clydach- all	87	4.23	0.58	0.43	0.28	<0.001
RU01	19	3.14	0.38	0.26	0.44	<0.001
RU02	22	3.99	0.41	0.31	0.31	<0.001
RU03	10	4.14	0.52	0.38	0.33	0.19
RU05	24	4.33	0.41	0.29	0.21	<0.001
RU06	21	4.46	0.47	0.32	0.33	<0.001
RU08	23	5.48	0.57	0.36	0.46	<0.001
Rudry- all	119	4.26	0.46	0.32	0.35	<0.001
DP01	22	5.24	0.62	0.37	0.49	<0.001
DP02	22	5.59	0.60	0.40	0.34	<0.001
DP03	19	6.02	0.67	0.52	0.30	<0.001
DP06	16	5.74	0.67	0.41	0.43	<0.001
Dinas- all	79	5.65	0.64	0.43	0.39	<0.001
all	580	4.59	0.55	0.39	0.33	<0.001

Table 4.5. Number of private alleles within Clydach sites

Site	Private alleles
$CL05^N$	3
$CL01^N$	5
CL12	8
CL02	1
CL04	1
CL11	4

Table 4.6 Pairwise Fst values between sites (with 9 loci) indicating population differentiation

	$CL05^N$	$CL01^N$	CL12	CL02	CL04	CL11	RU01	RU02	RU05	RU06	RU08	DP01	DP02	DP03	DP06
$CL05^N$	-														
$CL01^N$	0.09**	-													
CL12	0.02	0.05**	-												
CL02	0.05*	0.05**	0.03**	-											
CL04	0.07**	0.05**	0.02*	0	-										
CL11	0.06**	0.03**	0.03**	0.02*	0.02*	-									
RU01	0.2**	0.23**	0.21**	0.23**	0.26**	0.21**	-								
RU02	0.25**	0.26**	0.24**	0.27**	0.29**	0.24**	0	-							
RU05	0.27**	0.26**	0.24**	0.27**	0.29**	0.24**	0.04**	0.02**	-						
RU06	0.15**	0.20**	0.17**	0.18**	0.2**	0.17**	0.02**	0.03**	0.04**	-					
RU08	0.17**	0.17**	0.16**	0.15**	0.17**	0.14**	0.04**	0.03**	0.07**	0.05**	-				
DP01	0.13**	0.13**	0.09**	0.07**	0.08**	0.08**	0.18**	0.22**	0.22**	0.13**	0.14**	-			
DP02	0.09**	0.12**	0.07**	0.06**	0.08**	0.07**	0.14**	0.18**	0.18**	0.1**	0.11**	0.01	-		
DP03	0.06**	0.05**	0.03**	0.03**	0.06**	0.03**	0.15**	0.18**	0.18**	0.12**	0.1**	0.06**	0.04**	-	
DP06	0.09**	0.10**	0.08**	0.07**	0.1**	0.06**	0.06**	0.08**	0.08**	0.04**	0.04**	0.05**	0.04**	0.03**	-

*P <0.05

**P<0.01

Table 4.7 Pairwise Fst values between sites (with 8 loci) indicating population differentiation

	$CL05^N$	$CL01^N$	CL12	CL02	CL04	CL03	CL11	RU01	RU02	RU05	RU06	RU08	RU03	DP01	DP02	DP03	DP06
$CL05^N$	-																
$CL01^N$	0.10**	-															
CL12	0.04*	0.05**	-														
CL02	0.07**	0.05**	0.04**	-													
CL04	0.09	0.06**	0.03*	0.01**	-												
CL03	0.14**	0.13**	0.08**	0.16*	0.16**	-											
CL11	0.07**	0.03**	0.02**	0.02**	0.02**	0.12**	-										
RU01	0.32**	0.30**	0.28**	0.32**	0.35**	0.28**	0.28**	-									
RU02	0.27**	0.27**	0.25**	0.29**	0.32**	0.23**	0.24**	0.02**	-								
RU05	0.30**	0.29**	0.26**	0.29**	0.32**	0.26**	0.26**	0.02**	0.05*	-							
RU06	0.21**	0.26**	0.22**	0.25**	0.28**	0.23**	0.23**	0.04**	0.03*	0.04**	-						
RU08	0.18**	0.17**	0.16**	0.15**	0.18**	0.19**	0.13**	0.04**	0.04**	0.04**	0.05**	-					
RU03	0.33**	0.30**	0.28**	0.32**	0.34**	0.29**	0.27**	0.00**	0.03**	0.02**	0.06**	0.04**	-				
DP01	0.12**	0.12**	0.08**	0.06**	0.08**	0.16**	0.07**	0.23**	0.2	0.23*	0.18	0.12**	0.23**	-			
DP02	0.11**	0.13**	0.08**	0.07**	0.10**	0.13**	0.08**	0.19**	0.15**	0.19**	0.13**	0.10**	0.20**	0.00**	-		
DP03	0.07**	0.06**	0.04**	0.04**	0.08**	0.10**	0.03**	0.22**	0.19**	0.20**	0.19**	0.11**	0.21**	0.05**	0.05	-	
DP06	0.11**	0.09**	0.06**	0.04**	0.07**	0.15**	0.03**	0.22**	0.20**	0.21**	0.18**	0.10**	0.20**	0.03**	0.05**	0.02**	-

*P <0.05

**P<0.01

Table 4.8 Analysis of molecular variance without sites CL03 and RU03 (9 loci) and with sites CL03 and RU03 (8 loci)

	Source of variation	d.f.	Sum of squares	Variance components	%	P
9 loci	among regions	2	44.95	0.12	21.16	>0.001
	among sites within regions	12	17.86	0.03	5.57	>0.001
	among individuals within sites	535	213.11	0.4	73.27	>0.001
8 loci	among regions	2	44.15	0.1	19.81	>0.001
	among sites within regions	14	22.39	0.04	6.74	>0.001
	among individuals within sites	573	222.71	0.39	73.45	>0.001

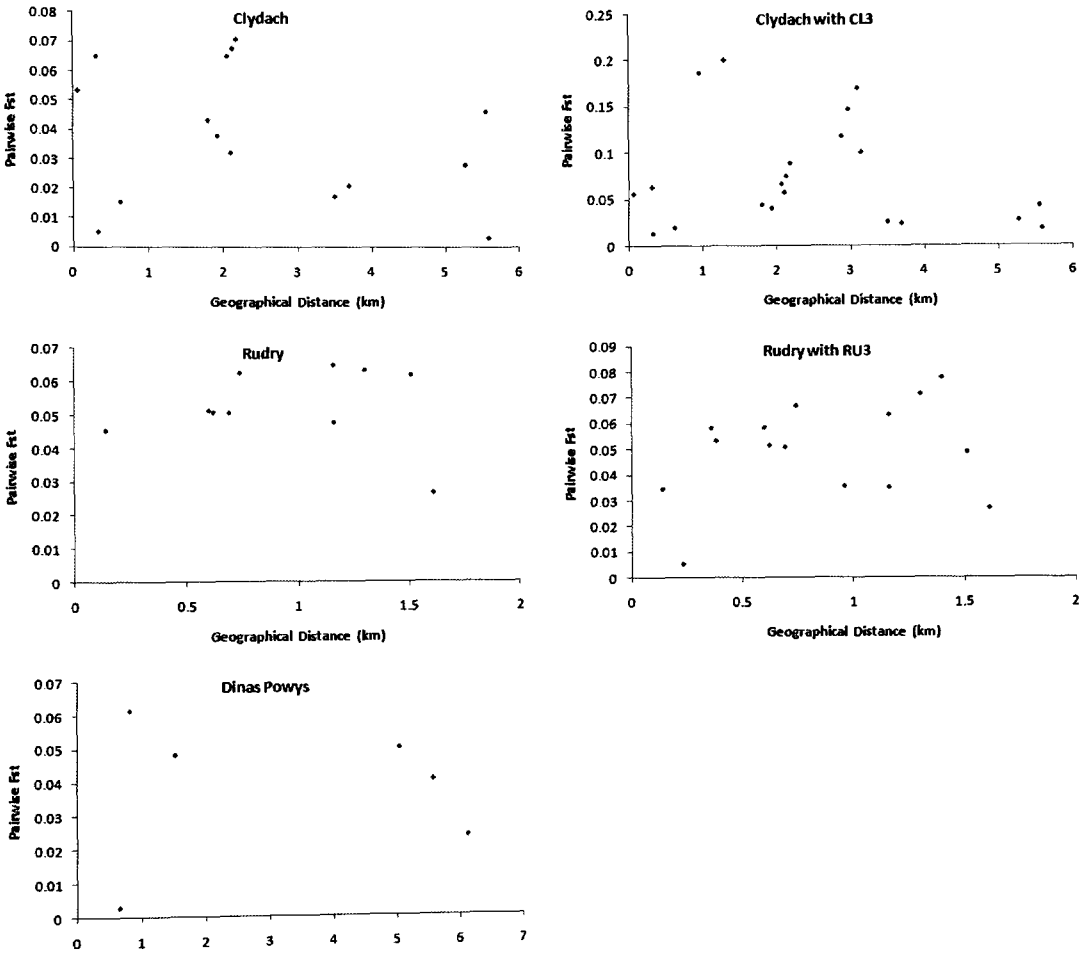
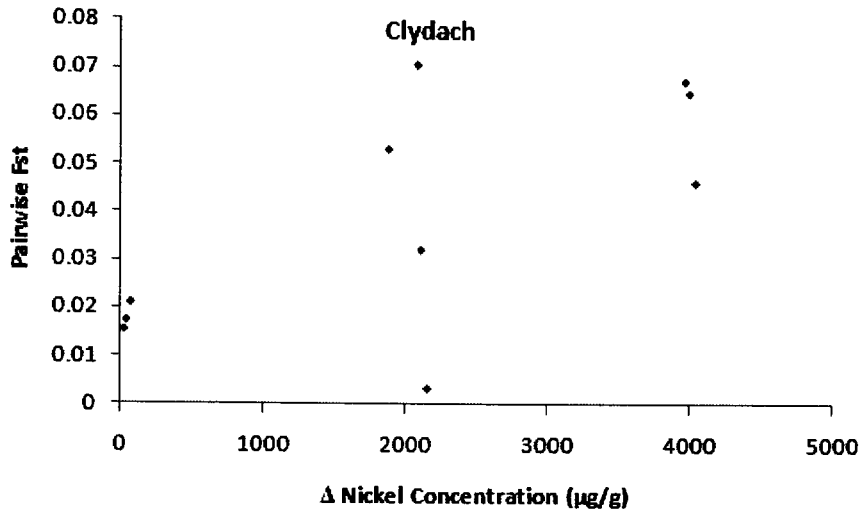


Figure 4.9 Correlations between pairwise Fst values and geographical distance



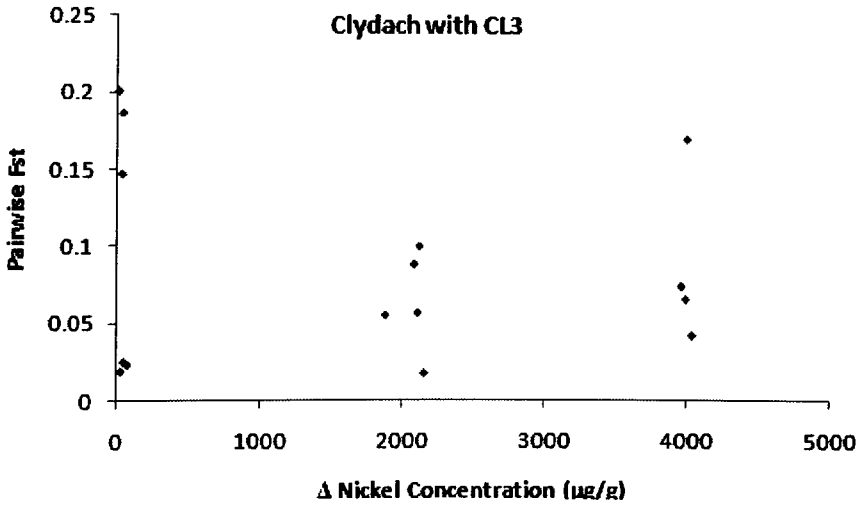
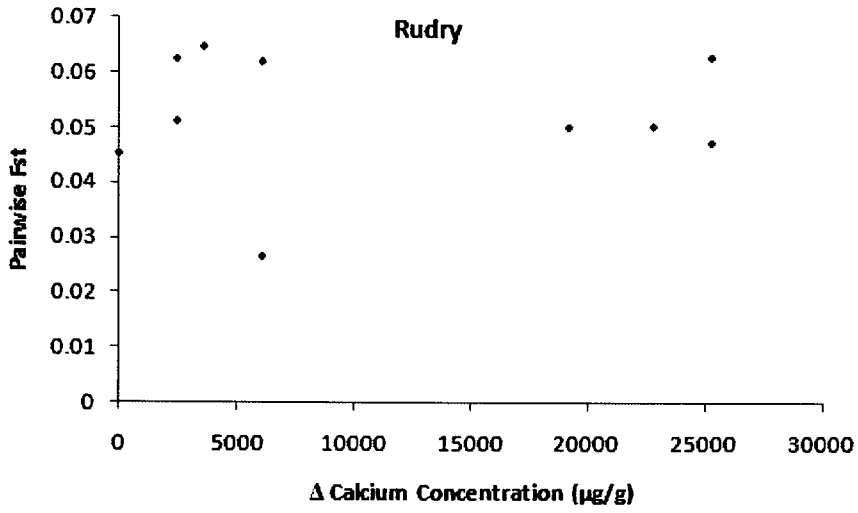


Figure 4.10 Correlation between pairwise Fst values and Δ soil nickel concentration at Clydach



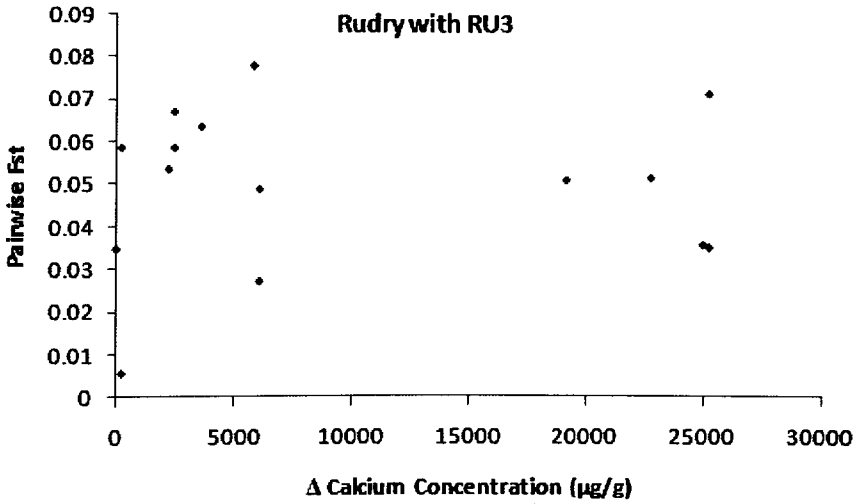


Figure 4.11 Correlation between pairwise Fst values and Δ soil calcium concentration at Rudry

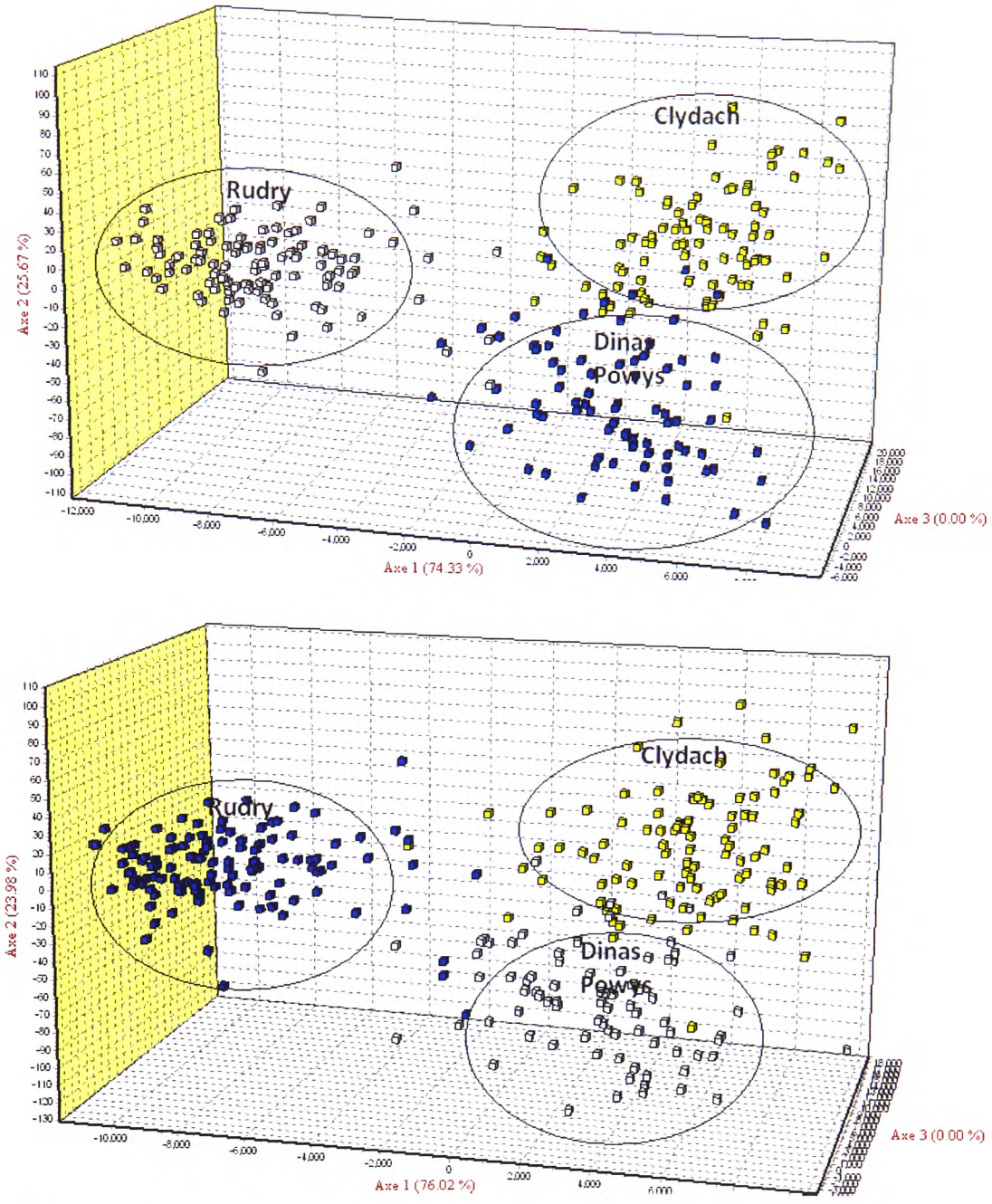


Figure 4.12 Three-dimensional FCA plots of populations from all three regions with 9 loci (above) and 8 loci (below)

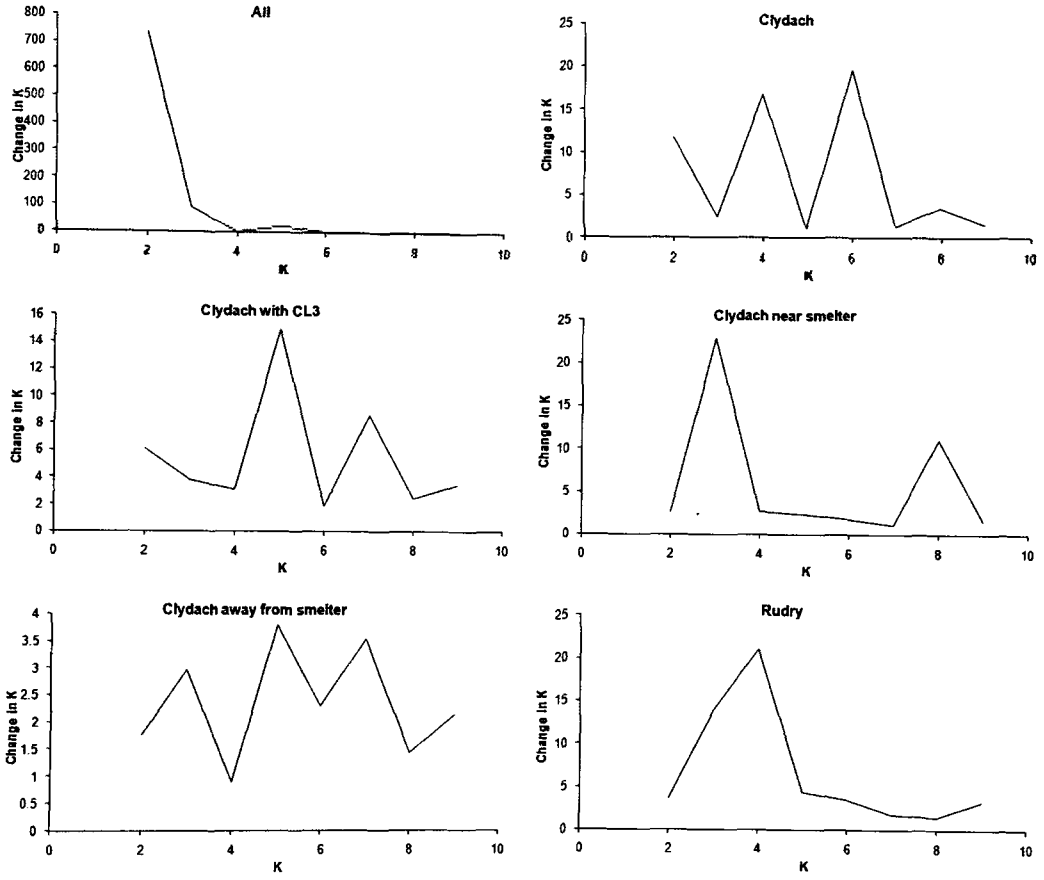


Figure 4.13 Determination of the optimal value of K (number of genetic clusters) from the preliminary Structure runs. This follows the method of Evanno et al. (2005) with the optimal value of K displaying the highest second order rate of change in log-probability value (designated Change in K on the graph)

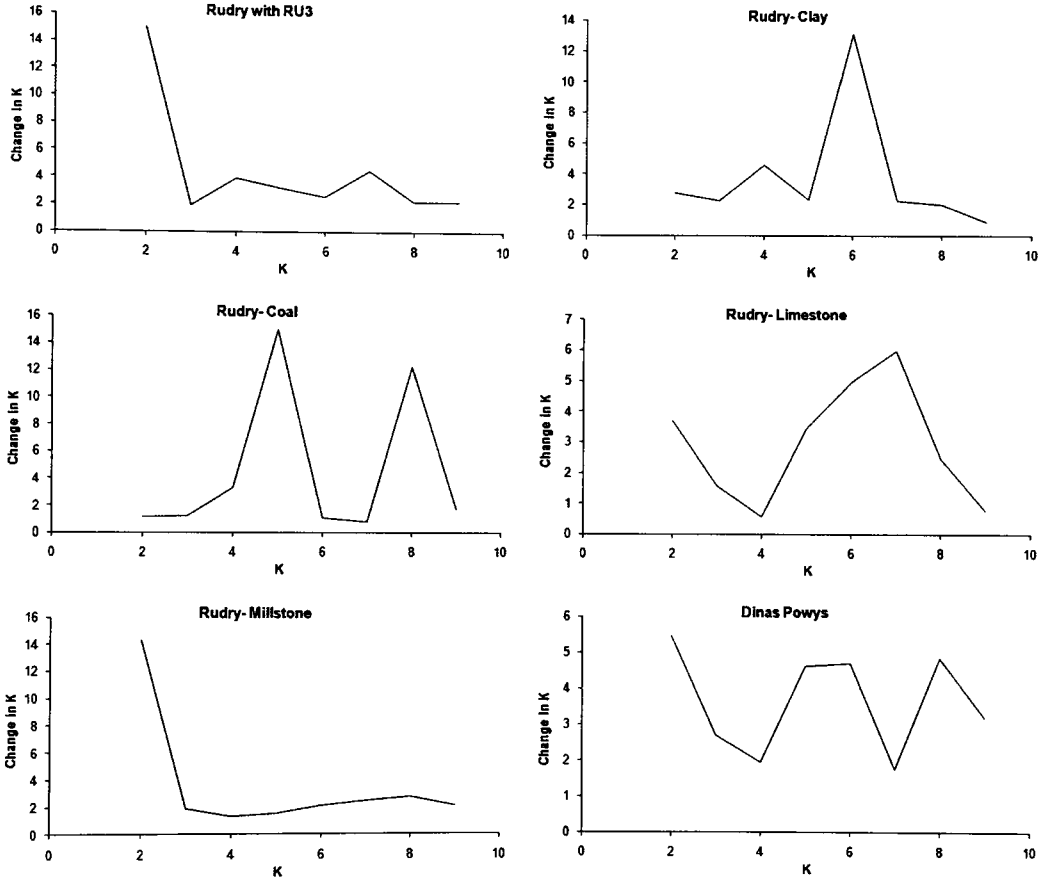


Figure 4.13 (continued) Determination of the optimal value of K (number of genetic clusters) from the preliminary Structure runs. This follows the method of Evanno et al. (2005) with the optimal value of K displaying the highest second order rate of change in log-probability value (designated Change in K on the graph)

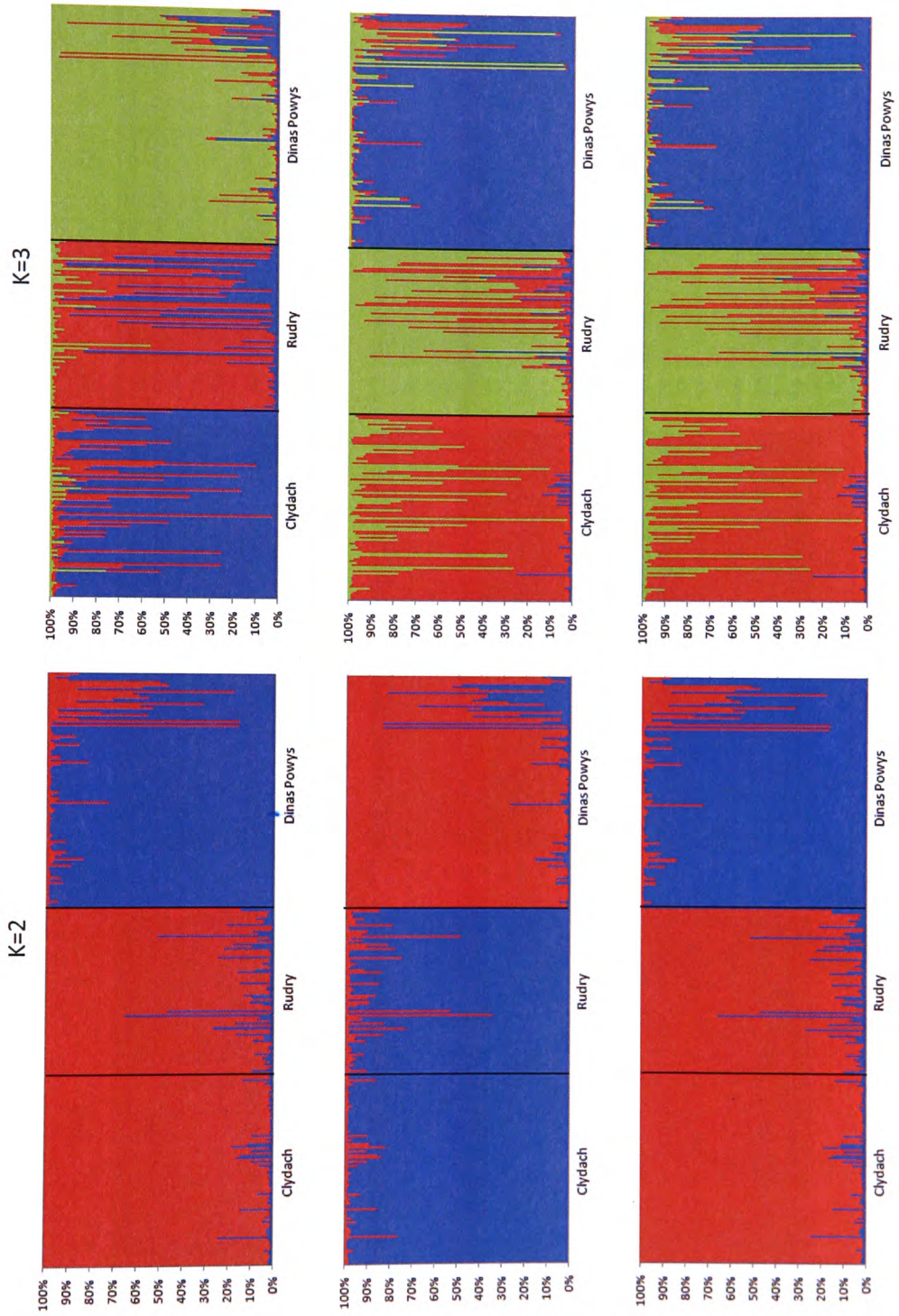


Figure 4.14 Results of Bayesian clustering analysis of all three regions using STRUCTURE. This analysis determines whether sampled 'populations' correspond to distinct genetic clusters (K=number of clusters in the analysis)

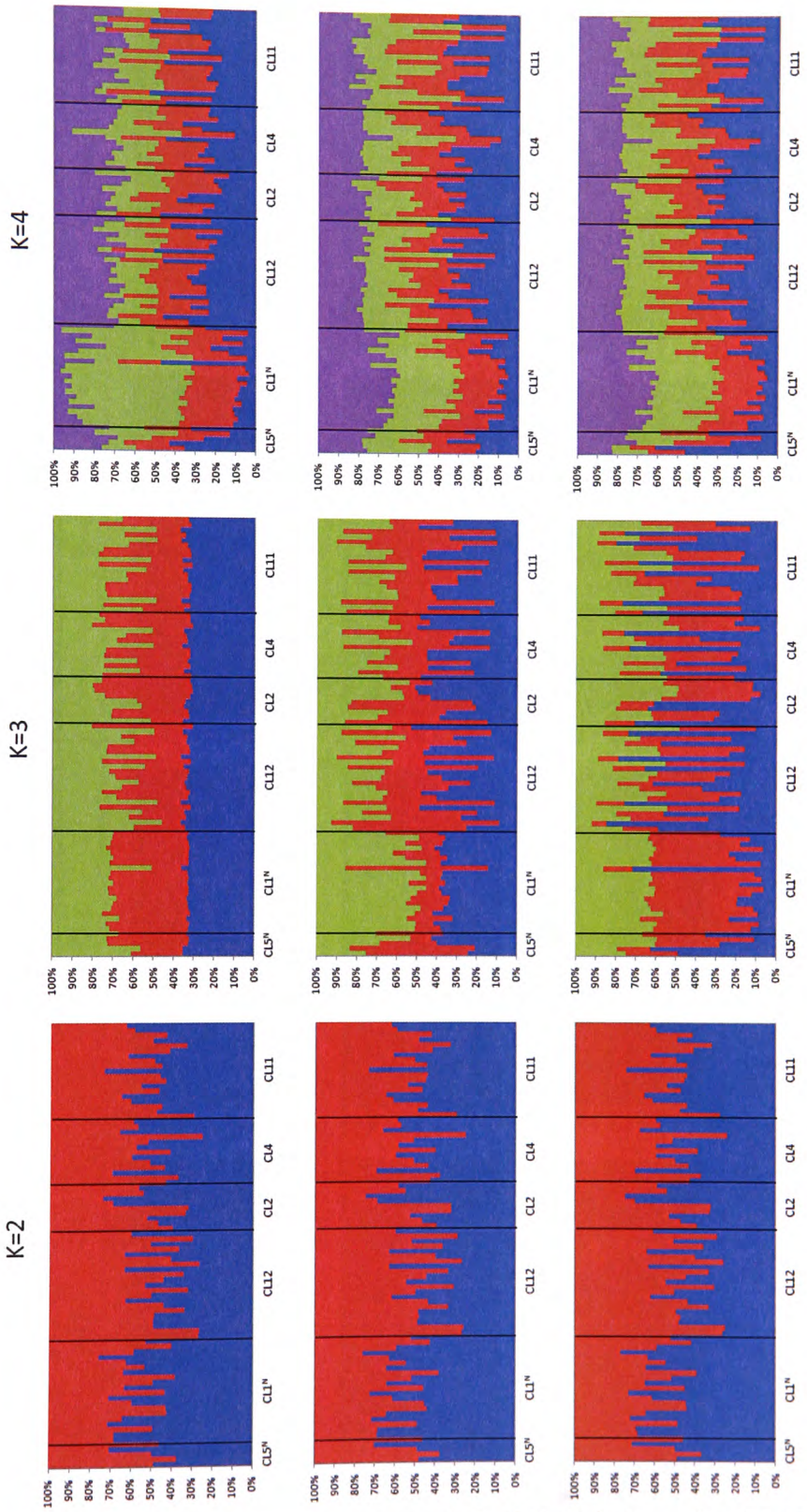


Figure 4.15 Results of Bayesian clustering analysis of Clydach sites using STRUCTURE. This analysis determines whether sampled 'populations' correspond to distinct genetic clusters (K=number of clusters in the analysis)

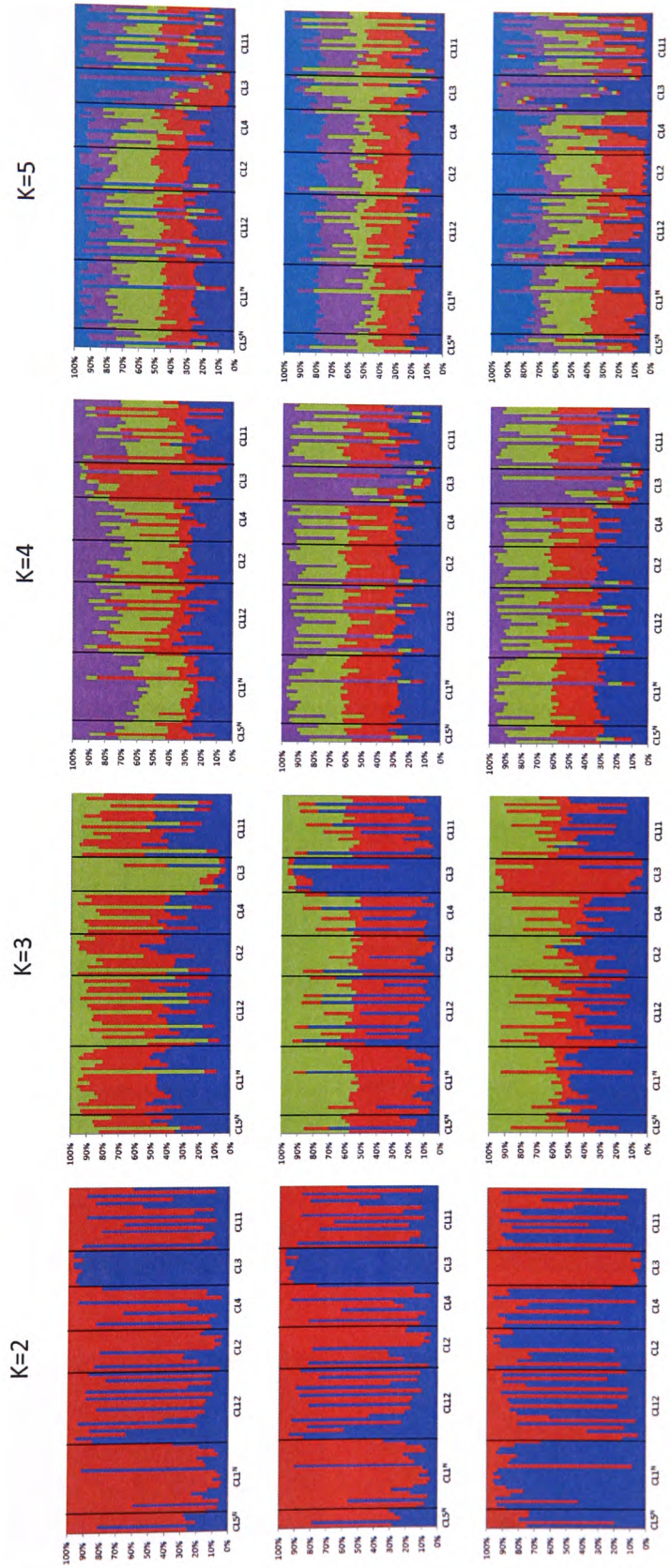


Figure 4.16 Results of Bayesian clustering analysis of Clydach sites (including CL3) using STRUCTURE. This analysis determines whether sampled 'populations' correspond to distinct genetic clusters (K=number of clusters in the analysis)

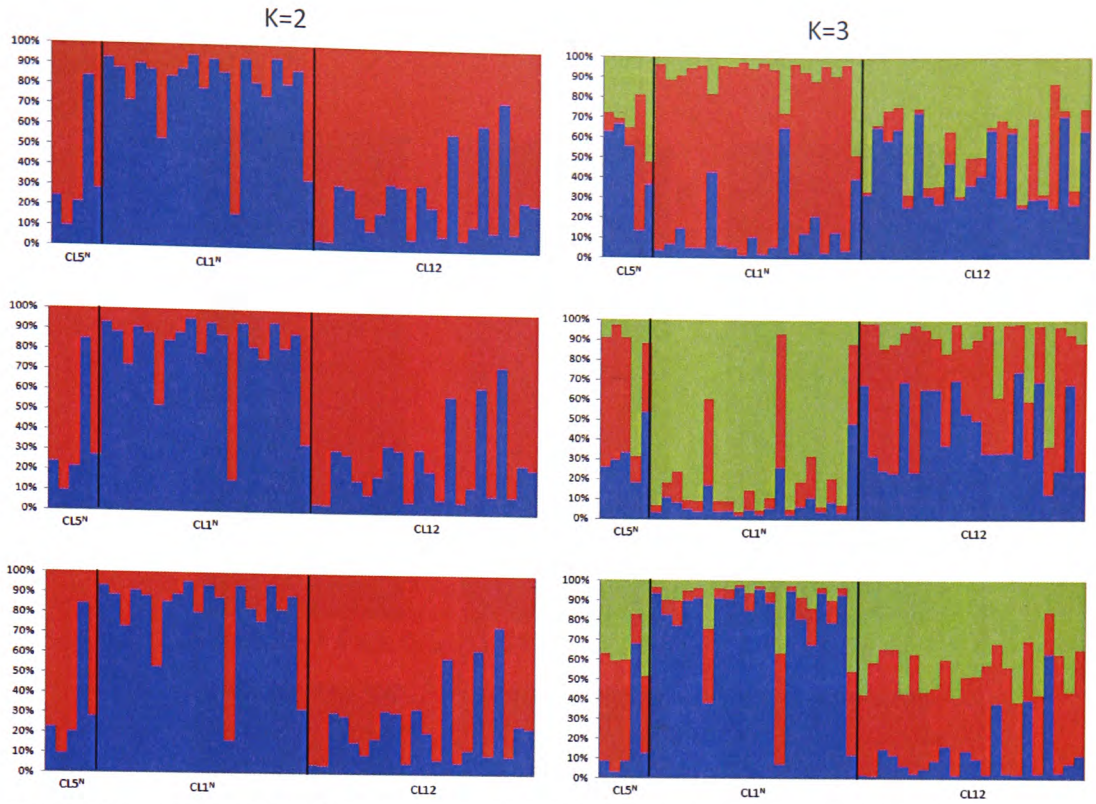


Figure 4.17 Results of Bayesian clustering analysis of Clydach sites near smelter using STRUCTURE. This analysis determines whether sampled 'populations' correspond to distinct genetic clusters (K=number of clusters in the analysis)

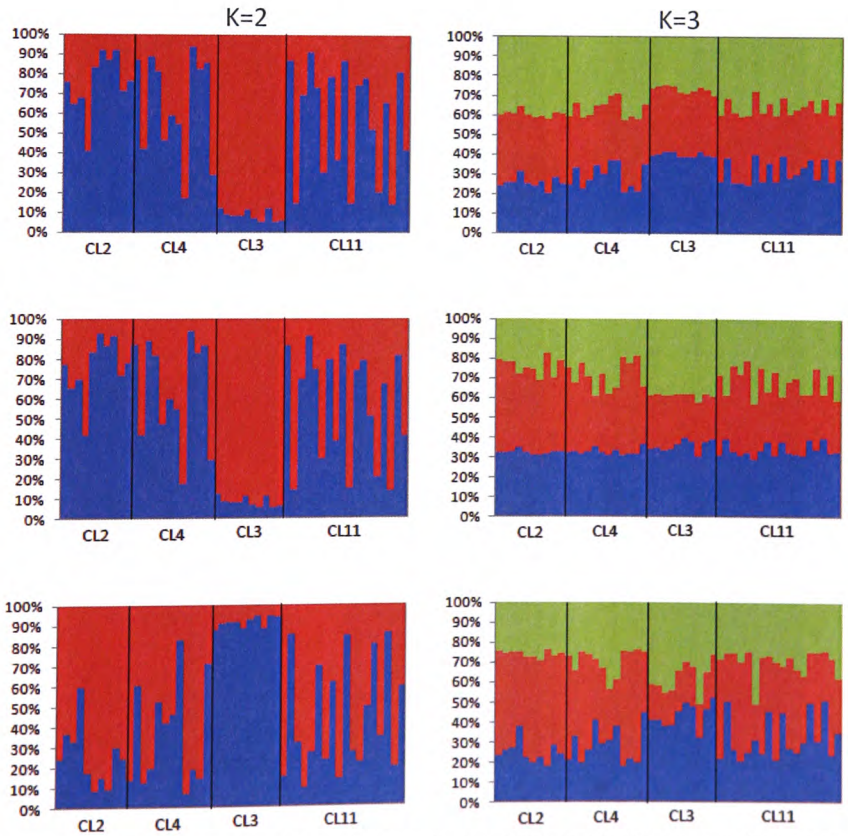


Figure 4.18 Results of Bayesian clustering analysis of Clydach sites away from smelter using STRUCTURE. This analysis determines whether sampled 'populations' correspond to distinct genetic clusters (K=number of clusters in the analysis)

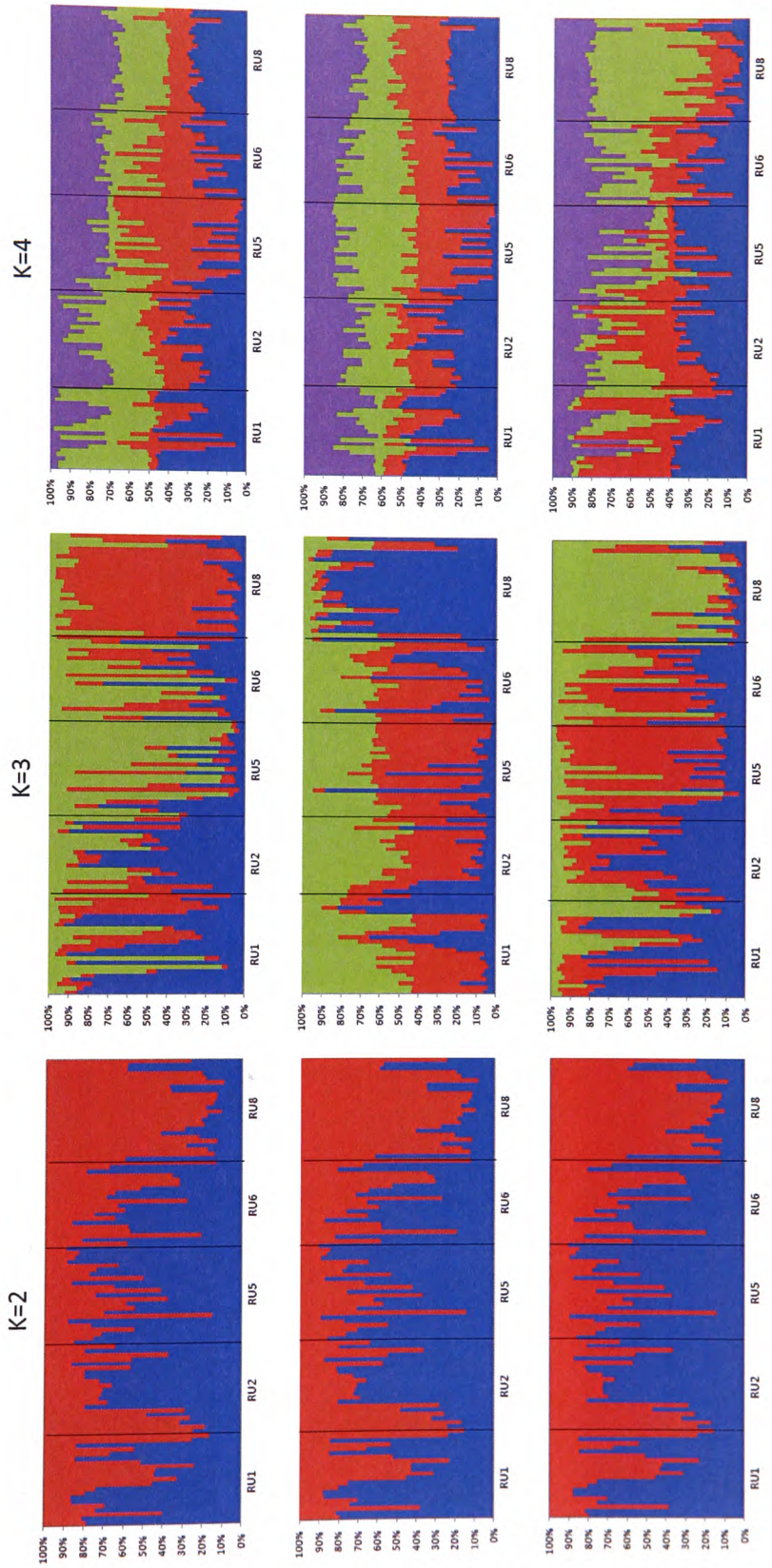


Figure 4.19 Results of Bayesian clustering analysis of Rudry sites using STRUCTURE. This analysis determines whether sampled 'populations' correspond to distinct genetic clusters (K=number of clusters in the analysis)

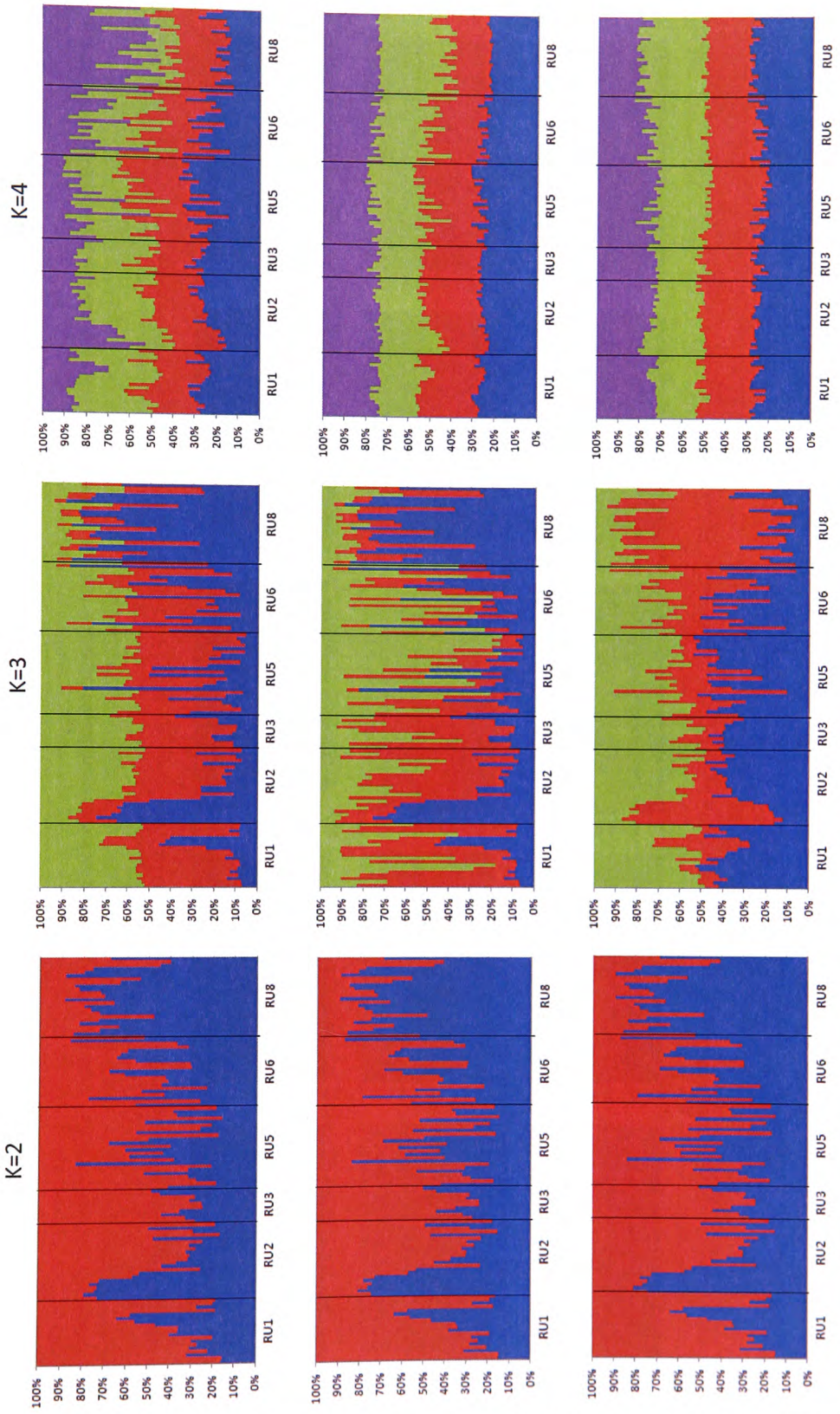


Figure 4.20 Results of Bayesian clustering analysis of Rudry sites (featuring RU3) using STRUCTURE. This analysis determines whether sampled 'populations' correspond to distinct genetic clusters (K=number of clusters in the analysis)

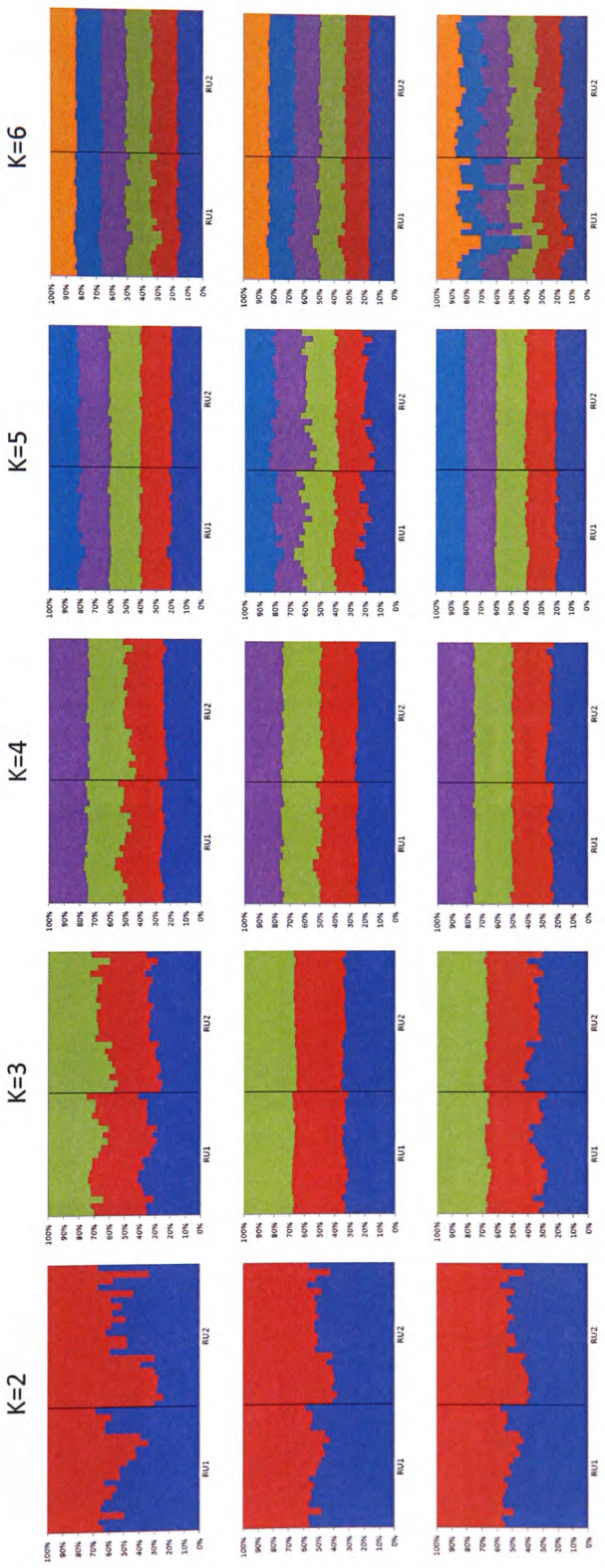


Figure 4.21 Results of Bayesian clustering analysis of clay Rudry sites using STRUCTURE. This analysis determines whether sampled 'populations' correspond to distinct genetic clusters (K=number of clusters in the analysis)

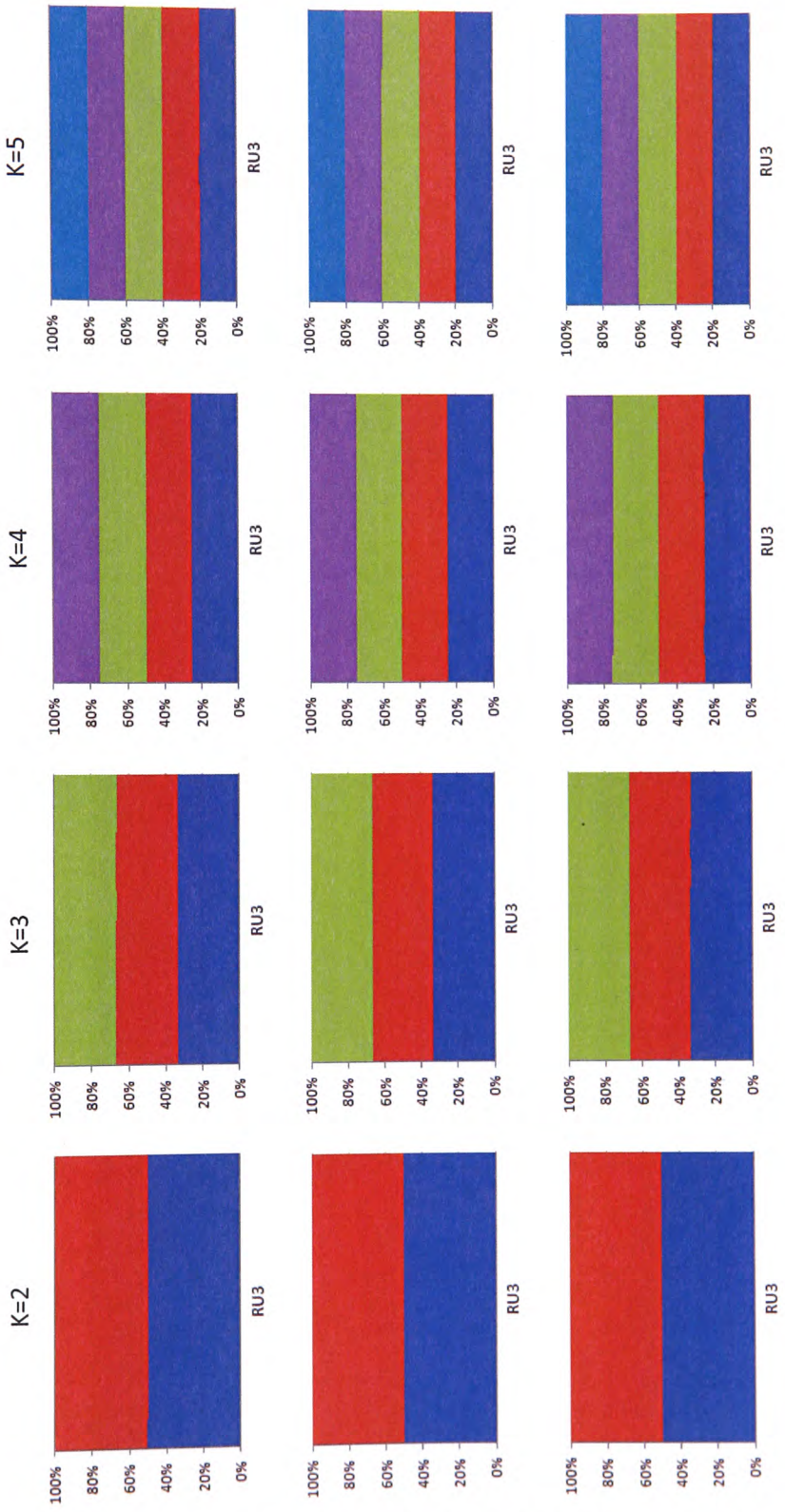


Figure 4.22 Results of Bayesian clustering analysis of coal Rudry sites using STRUCTURE. This analysis determines whether sampled 'populations' correspond to distinct genetic clusters (K=number of clusters in the analysis)

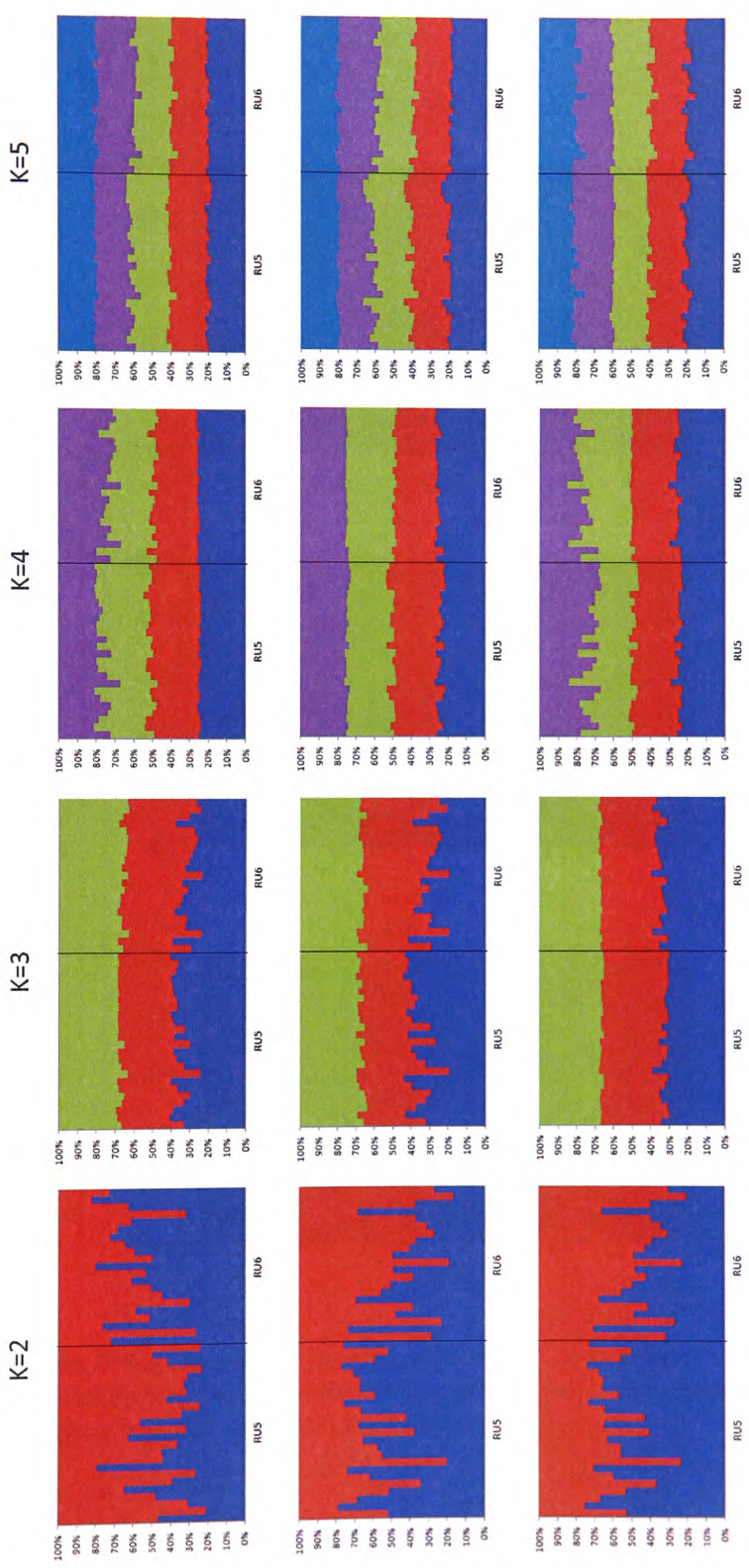


Figure 4.23 Results of Bayesian clustering analysis of limestone Rudry sites using STRUCTURE. This analysis determines whether sampled 'populations' correspond to distinct genetic clusters (K=number of clusters in the analysis)

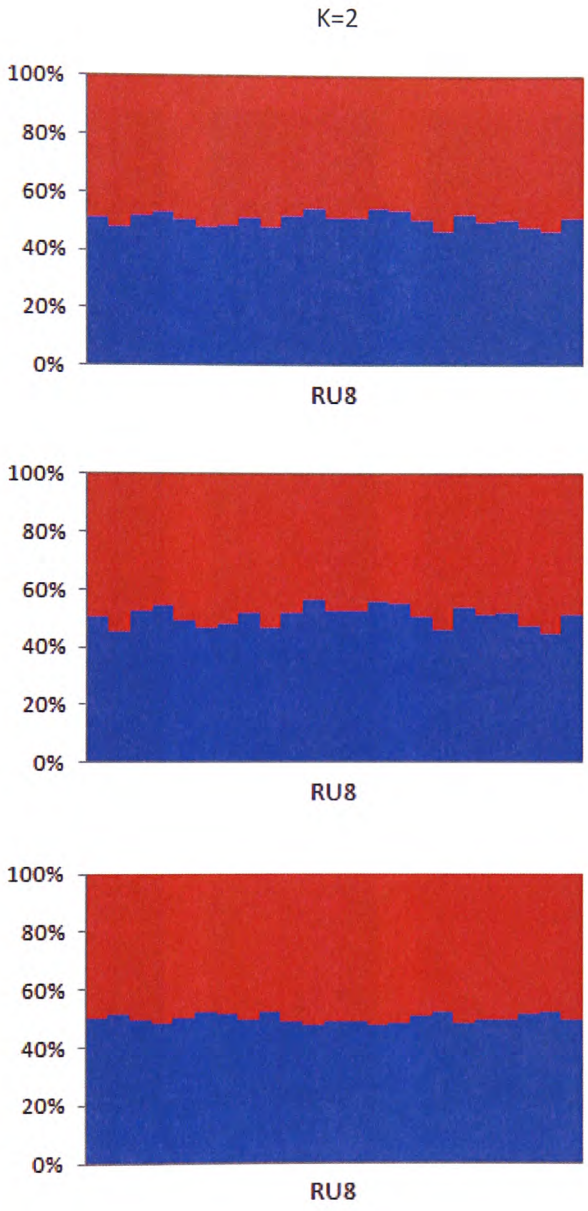


Figure 4.24 Results of Bayesian clustering analysis of millstone grit Rudry sites using STRUCTURE. This analysis determines whether sampled 'populations' correspond to distinct genetic clusters (K=number of clusters in the analysis)

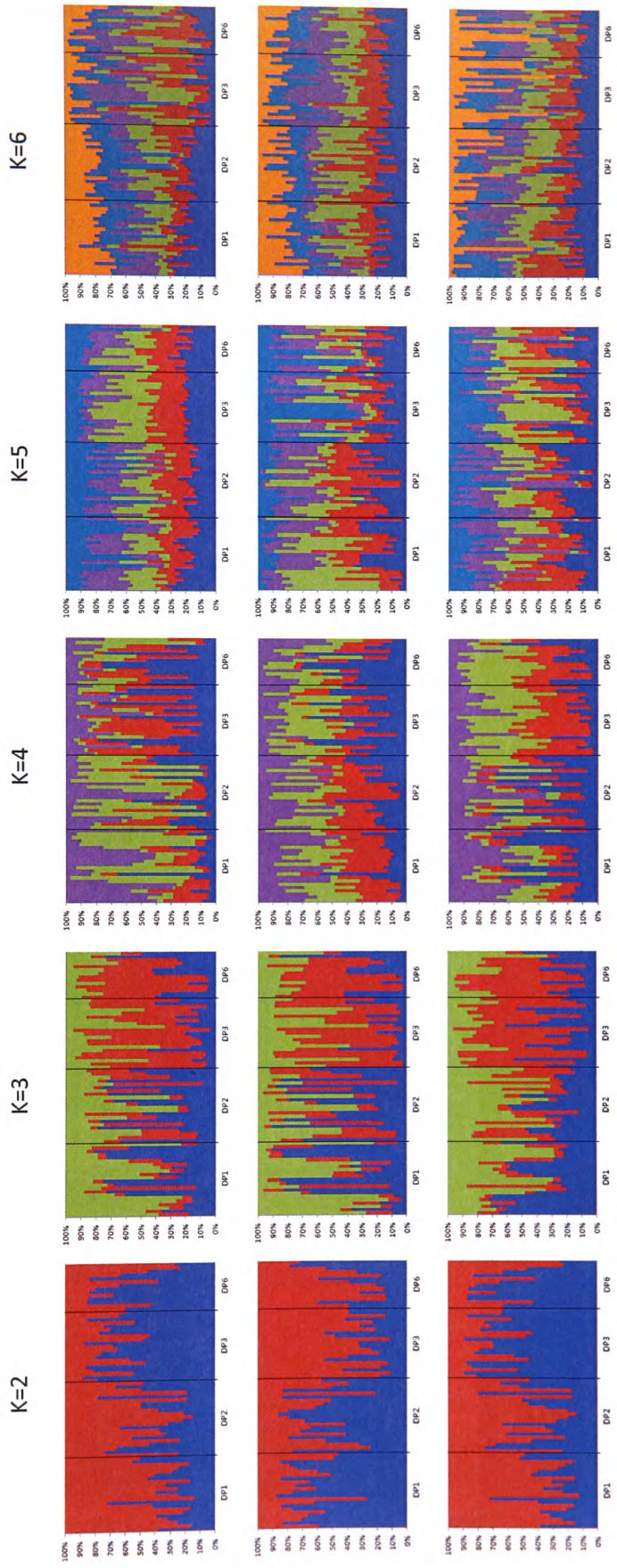


Figure 4.25 Results of Bayesian clustering analysis of Dinas Powys sites using STRUCTURE. This analysis determines whether sampled 'populations' correspond to distinct genetic clusters (K=number of clusters in the analysis)

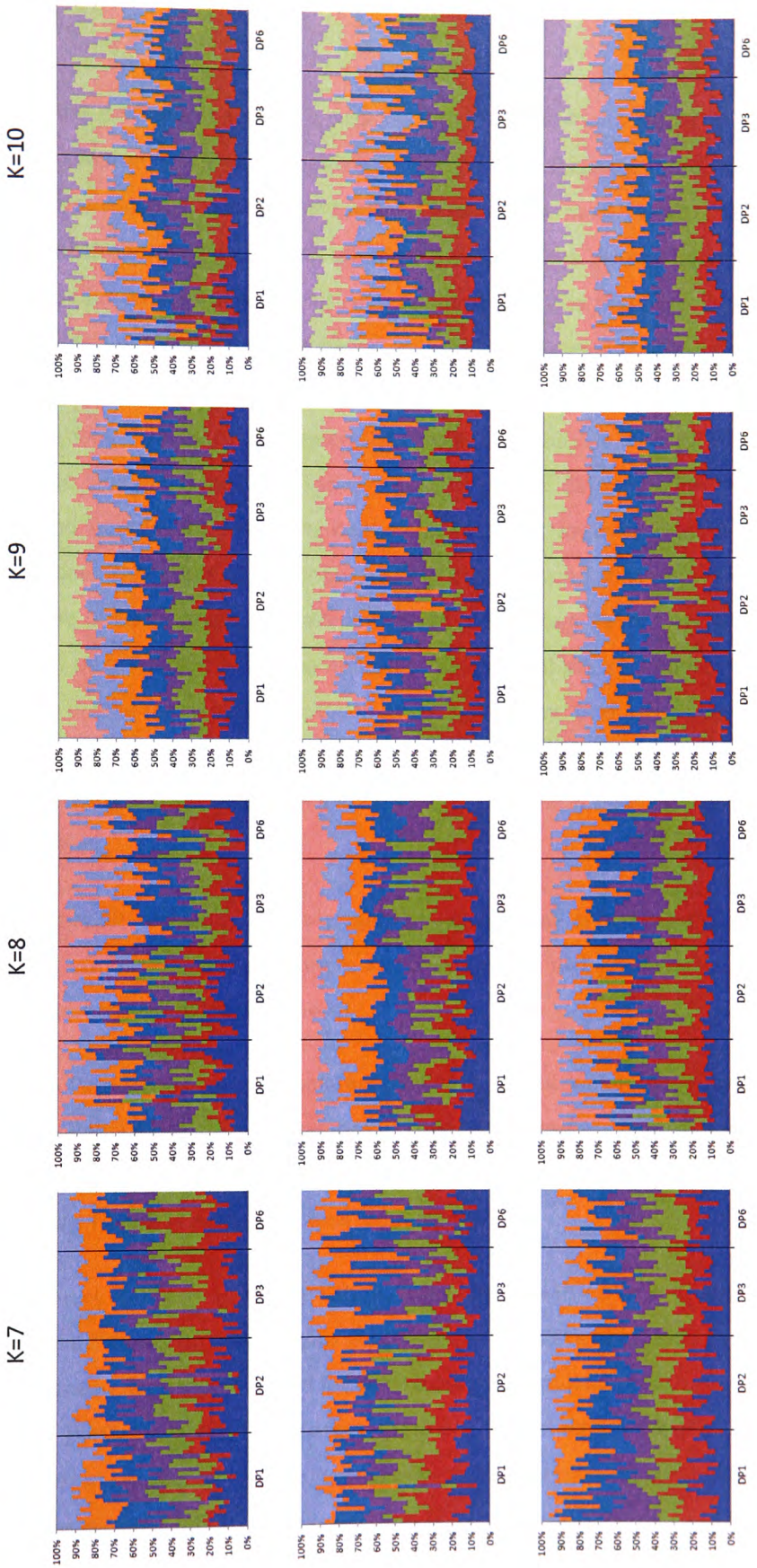


Figure 4.25 (continued) Results of Bayesian clustering analysis of Dinas Powys sites using STRUCTURE. This analysis determines whether sampled 'populations' correspond to distinct genetic clusters (K=number of clusters in the analysis)

Discussion

The ability of the earthworm species *L. rubellus* to yield meaningful results in ecotoxicological trials could be compromised if some populations are able to develop a genetically based tolerance towards the contaminant in question. If some populations are more resistant to a contaminant than others, uncertainties could arise in attempting to develop a general understanding of how sensitive *L. rubellus* is towards that particular contaminant. It is therefore important that any genetically adapted populations of *L. rubellus* are identified. Several studies have demonstrated that it may be possible to apply population genetics to infer the presence of genetically adapted *L. rubellus* populations at metal contaminated sites (Peles et al. 2003, Kille et al. 2010, Simonsen et al. 2010). These studies all indicated some degree of genetic differentiation between exposed populations and populations from less contaminated sites. However, one problem with this approach has been that some of the populations of contaminated sites have been exposed to a complex mixture of contaminants and environmental stresses (e.g. Peles et al. 2003, Simonsen et al. 2010). It can therefore sometimes be difficult to establish which contaminants or stressors are responsible for driving the genetic adaptation of the population. One way by which a correlation can be firmly established between the effects of selection and the concentration of individual metals is by studying population differentiation along a gradient of contamination (Posthuma and Van Straalen 1993).

To uncover possible evidence for the genetic adaptation of *L. rubellus* towards nickel contamination, populations were therefore sampled along a well-characterised nickel gradient enabling us to test the following hypotheses-

1. Populations are genetically adapted at the metal contaminated sites
2. Populations are phenotypically adapted at the metal contaminated sites
3. Populations at the metal contaminated sites are demographic 'sinks' sustained by gene flow from less exposed 'source' populations

Microsatellite analysis of *L. rubellus* populations from across all three regions uncovered evidence for a noticeable deficiency of heterozygous genotypes. All populations exhibited evidence of lower than expected mean heterozygoties and for the majority of populations this resulted in significant deviation from Hardy-Weinberg frequencies. There are several different explanations for such a low level of heterozygosity. Firstly, the non-amplification of certain microsatellite alleles due to mutations in the primer-binding region could have resulted in the occurrence of 'null alleles'. This has been observed in several studies, particularly those seeking to cross-amplify microsatellite loci using primers developed for a different species (Panova et al. 2008). Alternatively low heterozygosity can be detected within a population due to what is termed the 'Wahlund effect'

where several genetically differentiated subpopulations are unknowingly sampled as a single population. Finally, heterozygosity may be reduced within populations that are inbreeding, with the reproduction of closely related individuals resulting in a high degree of homozygosity among allele frequencies.

Of these possible explanations for a low degree of heterozygosity, a reduction in heterozygosity through the non-amplification of null alleles seems unlikely. The occurrence of null alleles is typically characterised by significant departures from Hardy-Weinberg equilibrium at one or more affected loci. However, the results of this study revealed departures from Hardy-Weinberg within seven of the nine microsatellite loci applied within the study. The loci that were affected also appeared to vary between populations in an inconsistent manner. Sequencing of microsatellite loci within both divergent lineages of *L. rubellus* has also revealed a high degree of sequence conservation within the flanking region of many of the microsatellites applied within this study (see chapter 3). This would be expected to limit variation within the primer-binding regions, reducing the opportunities for the development of null alleles.

The two remaining explanations for a low degree of heterozygosity: the presence of unrecognised sub-populations, i.e. Wahlund effect, and inbreeding, both seem like logical explanations given the sedentary nature of earthworm populations. It is possible that a Wahlund effect may have contributed to low heterozygosity values within some of the populations. Several other studies of earthworm populations have attempted to minimise the effects of substructuring by performing fine-scale sampling over microgeographic scales. Some of these studies have subsequently reported low levels of homozygosity (Velavan et al. 2009). Given the scarcity of *L. rubellus* individuals at many of the sampling sites in the present study, fine-scale sampling was often not possible and therefore earthworms were possibly sampled over a wider area than in previous earthworm studies. However, the genetic homogeneity of many of these sites indicated by cluster analysis would suggest an absence of widespread population sub-structuring.

An alternative explanation may therefore be that the sampled populations display a high degree of localised inbreeding. This has been suggested by other studies that have concluded that the sedentary nature of earthworm populations may result in a disproportionately high number of matings between closely related individuals (Novo et al. 2010a). It has been proposed that a combination of high reproductive rates and low dispersal rates within some epigeic earthworm species may result in them forming aggregations within field environments (Satchell 1955). The isolated nature of such aggregations could therefore result in a high degree of inbreeding.

Overall the pairwise values of genetic differentiation that were observed between populations within the three regions (Clydach, Rudry, Dinas) were largely comparable to the values stated by

Enckell et al. (1986) in their allozyme study of *L. rubellus* populations in the Faroe islands. This particular study featured populations from across a geographical scale that was comparable to the regions in the present study. The pairwise values of genetic differentiation between populations from the three different regions were however found to be far higher. An analysis of molecular variance in the present study indicated that the proportion of genetic variance partitioned between regions exceeded that found within populations. Populations from the three study regions were also clearly differentiated by Bayesian cluster analysis.

The degree of genetic differentiation found between sites from the three regions did however appear to be counter-intuitive in that it was not correlated with the geographical distances between the three regions. As an example of this, sites located within the region of Dinas Powys appeared to be more highly differentiated from the geographically closer sites of Rudry than they were from the sites of Clydach.

Further anomalous patterns of genetic differentiation were uncovered within some of the regions with some sites appearing highly differentiated from others within close proximity. As an example of this within Clydach the population of site CL03 was found to be highly differentiated from the other sites and genetically closer to some of the sites within Dinas Powys.

Comparable incidences of differentiation were uncovered in the allozyme study of Shepeleva et al. (2008) who uncovered geographically irregular patterns of differentiation between populations of *L. rubellus*. One possible explanation for these anomalies could be the passive dispersal of earthworms between different geographical areas. Studies have indicated that human agricultural activities could play a major role in dispersing *L. rubellus* individuals (Marinissen and van den Bosch 1992). Given that *L. rubellus* is an epigeic species (Sims and Gerard 1999), residing close to the soil surface, the species could be pre-disposed to such passive dispersal through the movement of soil.

Alternatively geographical patterns of differentiation between *L. rubellus* populations may have been affected by the cultivation of field sites. The cultivation of soils at these sites may have imposed selective pressures upon earthworms. This may explain why the region of Rudry was indicated to be the most highly differentiated region of the three as it featured sites from non-cultivated land. The study of Enckell et al. (1986) indicated that selective pressures from agricultural activity may constrain genetic divergence between *L. rubellus* populations. In this study earthworm populations inhabiting historically cultivated sites were found to be genetically closer to one another than they were to geographically proximate populations from non-cultivated sites.

Microsatellite analysis of the Clydach sites did not provide strong support in favour of the hypothesis that populations at the most heavily nickel contaminated sites are sustained by the local recruitment of adapted individuals. This was in spite of the fact that soil nickel concentrations at some of the sites closest to the smelter were found to be far higher than levels that have been demonstrated to affect the survival and reproduction of *L. rubellus* (Lister et al. 2011). Pairwise Fst values indicated a low to moderate degree of genetic differentiation to exist between the Clydach sites. Although the mean pairwise Fst values for two of the sites closest to the smelter ($CL01^N$ and $CL05^N$) did appear to be slightly higher than those of the other Clydach sites this was a marginal effect and it is unclear if this was due to the genetic adaptation of *L. rubellus*. Neither of the statistical tests performed was able to uncover a significant correlation between the soil nickel concentration and the genetic differentiation of the Clydach sites. It is also worth mentioning that the mean pairwise Fst value for the site of CL12 which was located under 1km from the smelter was closer to the values obtained for the less contaminated sites of Clydach and the two control regions. Unfortunately as a soil sample was not obtained for this site it is currently unclear whether the nickel contamination at this site is comparable to that of $CL01^N$ and $CL05^N$, but this is likely to be the case. The degree of genetic differentiation that exists between both $CL01^N$ and $CL05^N$, and the other sites is also not particularly high when it is compared with the differentiation that has been uncovered between populations in existing microsatellite studies of genetic adaption (Plath et al. 2007).

The results of Bayesian clustering analysis indicated a considerable degree of admixture to exist between all of the sites within Clydach except for the previously mentioned highly differentiated site of CL03. $CL01^N$ was however found to form a distinct cluster during an analysis featuring individuals from the three sites closest to the smelter ($CL01^N$, $CL05^N$, CL12). The urban nature of this site however makes it unclear whether this is an effect of nickel contamination or of other anthropogenic factors.

It is possible that the level of genome-coverage provided by microsatellite markers may have been insufficient to uncover evidence of adaptation if selection is weak. Microsatellites are often used to infer the adaptation of populations as they can become differentiated through genetic drift following a selective sweep (Simões et al. 2008). However if adaptation has occurred with a relatively weak level of selection, high levels of gene flow or sexual reproduction could have maintained the neutral genetic diversity of these populations (Muller et al. 2007). Several incidences of adaptation have been uncovered where selection has not resulted in a strong selective sweep meaning that only a small number of loci are genetically differentiated (e.g. McMillan et al. 2006). In such incidences multi-locus genetic markers have been more successful in uncovering evidence of adaptation. An analysis of the populations within the present study has

however been conducted using multi-locus AFLP (amplified fragment length polymorphism) markers. This study failed to uncover any evidence for the increased genetic differentiation of nickel-exposed populations (G. Juma personal communication) although analysis for signatures of genomic selection is ongoing.

Despite the existence of possible limitations in the genome coverage offered by microsatellites, they should still be capable of uncovering evidence of genetic bottlenecks given that such events would change gene frequencies at loci throughout the genome. The fact that no such evidence has been uncovered within the present study indicates that the high degree of nickel contamination at Clydach may not had little effect upon the demography of the resident *L. rubellus*. There are however several possible explanations as to why the populations of *L. rubellus* inhabiting the most heavily nickel-contaminated sites of Clydach do not appear to display the kind of elevated genetic differentiation that has been uncovered in comparable population genetic studies (e.g., Peles et al. 2003, Andre et al. 2010, Simonsen et al. 2010).

One explanation for the lack of differentiation may be that populations of the most heavily contaminated sites may actually represent sink populations sustained by the migration of non-resistant individuals from less contaminated sites. This explanation is supported by the finding that nickel concentrations at some of the more contaminated sites are higher than levels that have been demonstrated to cause toxic effects within *L. rubellus* during exposure trials (Lister et al. 2011). Although the imperfect nature of the microsatellites (see chapter 3) limited the analysis methods available as a stepwise mutation model could not be assumed, the results that were obtained did not display any clear signatures of sink populations. The values of genetic diversity for sites closest to the smelter were not found to be higher those for sites located further away. The average values of genetic differentiation for sites closest to the smelter were also not found to be markedly lower than those for sites located further away. Both of these findings contradict the expected patterns of genetic diversity and differentiation for sites featuring a large number of immigrant individuals (i.e. sink populations). In addition, although the number of private alleles found at sites near the smelter did appear to be slightly higher than some of the sites further away, it is unclear whether this is indicative of an increased number of immigrants within these sites or alternatively of gene mutation following exposure to nickel. Evidence of a high degree of genetic admixture was also uncovered throughout Clydach during Bayesian cluster analysis. It was therefore not possible to discern whether the two most contaminated sites displayed a disproportionally high degree of immigrant individuals from this analysis.

A second explanation for the presence of *L. rubellus* at the most contaminated sites may be that these populations are able to tolerate the high levels of nickel through a process of phenotypic plasiticity rather than by genetic adaptation (Morgan et al. 2007). Such phenotypic plasticity could

include the physiological acclimation of adult individuals or the developmental acclimation of juveniles (Meyers and Bull 2002). It has been proposed that phenotypic plasticity may occur within a stressed population as a precursor to the development of full genetic adaptation (Crispo 2008). Alternatively a population may display tolerance through phenotypic adaptation to a contaminant instead of adaptation if there are high fitness costs associated with the expression of genetically tolerant genotypes (Posthuma and Van Straalen 1993, Morgan et al. 2007). This could arise due to high metabolic costs of maintaining tolerance mechanisms (Pook et al. 2009) or due to pleiotropic effects.

One final possibility is that the bioavailability of nickel to *L. rubellus* may have been low at Clydach despite the high total soil concentration. This could have occurred if the soil concentration of nickel was sequestered into a largely immobile form reducing its uptake by earthworms (Voua Otomo 2011). Additionally physical factors such as soil pH could have acted to reduce the uptake of nickel in pore water (Peijnenburg et al 1999). The processing of samples for metabolomic analysis meant that it was not possible to determine body concentrations of nickel, which might have enabled a determination of the availability of nickel at the most contaminated sites. It was also decided not to apply soil pH (in the form of calcium concentration) as a cross-factor during population analyses as pH data was only available for two of the heavily contaminated sites.

In conclusion the results of the present study do not support the presence of genetically adapted *L. rubellus* individuals at the most heavily nickel-contaminated sites of Clydach. Although there was some evidence of a slightly higher level of genetic differentiation existing between two sites close to the smelter and those further away, this was not of a level that would indicate the occurrence of strong selection. Other microsatellite studies of adapted populations have uncovered levels of genetic differentiation between geographically proximate adapted and non-adapted populations that are comparable to that between populations from different geographical areas (Plath et al. 2007). Such a high degree of differentiation was not evident in this study. If selection for a genetically-based metal-tolerance has occurred at the most heavily contaminated sites of Clydach, it must therefore have been relatively weak. The existence of weak selection within these sites would however seem to contradict the findings of the metal-analysis which indicated the nickel contamination at these sites to be of a level that far exceeds that known to increase the mortality of *L. rubellus* (Lister et al. 2011). It is therefore possible that *L. rubellus* at the most contaminated sites of Clydach may display tolerance of nickel contamination through a form of phenotypic plasticity (Beldade et al. 2011). Although no evidence of nickel-tolerance through phenotypic plasticity was uncovered in the exposure-trial of Lister et al. (2011), their study only focused upon the effects of nickel upon adult individuals. It is therefore possible that *L. rubellus* individuals exposed to nickel may be able to become more tolerant by another

unforeseen form of plasticity such as developmental acclimation (Meyers and Bull 2002).

Chapter 5. General Discussion

Background to study

It has been proposed that taxa that are applied as sentinel species should display a minimal amount of genetic variability (Hilty and Merenlender 2000). The application of this criterion to ecotoxicological studies aims to reduce the influence of possible differences in tolerance between sentinel populations that could mask the true degree of environmental stress within that environment. Two possible sources of genetic heterogeneity within a sentinel species however are the existence of unrecognised cryptic genetic lineages and the existence of populations that have genetically adapted to the environmental stress under investigation.

The increasing application of genetic analysis techniques to the identification of species has uncovered evidence of divergent cryptic lineages within many sentinel species (e.g. Chenon et al. 2000, Wares et al. 2007). Further studies of some cryptic lineages in exposure trials have subsequently revealed them to differ markedly in their tolerance towards contaminants (Sturmbauer et al. 1999, Linke-Gamenick et al. 2000).

Examples of genetic adaptation have also been found within populations of many important ecotoxicological sentinels (e.g. Klerks 2002, Diogo et al. 2007, Langdon 2009). The individuals of some of these populations have demonstrated a genetically-based resistance towards contaminants by producing resistant F1 and F2 offspring during laboratory breeding studies (Langdon et al. 2009). Studies have also indicated that selection may drive the rapid development of genetic adaptation within some exposed populations (Klerks and Levinton 1989, Spurgeon and Hopkin 2009).

Studies indicate that the widely applied earthworm sentinel *Lumbricus rubellus*, may display both of the aforementioned forms of genetic heterogeneity within its populations. The studies of King et al. (2008) and Andre et al. (2010) have both indicated the existence of divergent cryptic lineages within populations of *L. rubellus*. These studies have also demonstrated the existence of sympatric populations of the two lineages and the presence of these lineages at contaminated sites which could present a problem for ecotoxicological studies. In order to develop a deeper understanding of the degree of divergence between these two lineages they were analysed in two separate studies. The first of the studies featured a wider mitochondrial sequence analysis of the two lineages from populations located throughout southern England and Wales. A morphological analysis was also conducted to uncover any possible lineage-specific traits that could be applied to differentiate the individuals of these two lineages. The second study applied microsatellite fragment and sequence analysis to investigate the degree of nuclear genetic differentiation

between the lineages. This would uncover any evidence of male-mediated introgression between the lineages.

Several studies have also indicated that some metal-exposed populations of *L. rubellus* may display a genetically based resistance towards contaminants. These studies include laboratory based exposure trials featuring F1 and F2 offspring (Langdon et al. 2001, Langdon et al. 2009) and population genetic studies (Andre et al. 2010, Simonsen and Klok 2010). In order to observe whether *L. rubellus* populations are genetically adapted to nickel contamination a population genetic study was conducted using microsatellite markers. Populations were sampled along a gradient of nickel contamination enabling patterns of genetic differentiation to be observed. This would allow the inference of genetically adapted populations.

Synthesis of the findings

The mtDNA COI analysis within the present study indicated the two lineages of *L. rubellus* to be even more divergent than previous studies, with a mean uncorrected p-distance of 14% between these two lineages. This is far higher than the value stated by King et al. (2008) of 8.28% and is only slightly lower than the mean level of divergence uncovered between *L. rubellus* and the congeneric species, *L. festivus*.

MtDNA analysis also uncovered a large number of new haplotypes within both of the *L. rubellus* lineages. Within the two lineages levels of sequence divergence between haplotypes were found to be lower within lineage B. The halpotypes of lineage B differed by up to 6bp which contrasted with lineage A where some of the haplotypes were found to differ by 17bp. The relatively lower level of genetic divergence observed between the haplotypes of lineage B could therefore indicate that these haplotypes emerged more recently than many of those found within lineage A.

The results of the microsatellite analysis also indicated a strong differentiation of the two lineages. The differentiation was apparent both in terms of allele frequencies within the lineages and in the degree of sequence variation revealed during alignment of flanking and repeat regions. Allele frequency measurements indicated that the degree of genetic differentiation between sympatric populations of the two lineages was far greater than the differentiation between allopatric populations of the same lineage. This differentiation was also reflected in both the results of Bayesian cluster analyses using NEW-HYBRIDS and STRUCTURE, and a principal coordinates analysis that all clearly differentiated the lineages within the two sites where they co-occurred. None of the cluster analyses was however able to uncover the existence of hybrid individuals, as were found in the study of Andre et al. (2010).

The finding that microsatellite data can successfully differentiate the two lineages of *L. rubellus* mirrors the findings of Dupont et al. (2011) who were able to differentiate several of the mtDNA lineages uncovered by King et al. (2008) using microsatellites. Some of the findings of Dupont et al. (2011) were also found to be congruent with inferences of nuclear differentiation uncovered by King et al. (2008) using amplified fragment length polymorphism data. This therefore indicates that microsatellites may be comparable to AFLPs and other nuclear molecular markers in their ability to differentiate individual genetic lineages or species.

Given that no evidence of hybridisation was uncovered in the current study it is unclear whether the hybridisation indicated in the AFLP study of Andre et al. (2010) is a frequent occurrence between the lineages of *L. rubellus* or a rare event. The study of Andre et al. (2010) did feature twice as many populations than the present study however, although two of these populations were predominantly composed of a single lineage. All of the populations were also located within a single geographical region that was contaminated with lead. One possible explanation for the existence of this hybridisation may therefore be that it is a result of selection imposed by lead contamination. It has been proposed that hybrid individuals may sometimes display an elevated degree of fitness or 'hybrid vigour' compared with non-hybridising individuals (Burke and Arnold 2001).

Alternatively dissimilarity between the results of microsatellite and AFLP analysis could reflect differences in the accuracy of these two molecular markers. Although the wide-ranging genomic coverage of AFLP markers is useful for uncovering hybrid individuals, AFLP loci can often exhibit size homoplasy (Gort and Eeuwijk 2011). This can cause separate loci to be mistakenly grouped together as a single locus, leading to over-estimation of the degree of relatedness between individuals (Gort and Eeuwijk 2011). It is therefore possible that hybrid individuals identified during AFLP studies of *L. rubellus* may have resulted from the scoring of homoplasic loci.

A morphological analysis of the two lineages uncovered an anteriorally located trait that appeared to differ between the two lineages and could therefore represent a possible identifer of these lineages. The trait which appeared to represent a glandular tumescence, was tested in a blind-trial and found to display success in identifying lineage B individuals. A means of accurately identifying lineage A individuals was also suggested by the finding that this trait was only found to be expressed over more than one segment within this particular lineage. No mention of this morphological trait could be found within the taxonomic guide of Sims and Gerard (1999). However the taxonomic description of *L. rubellus* by Bouché (1972) mentions an anteriorally located glandular area that appears to correspond to this trait.

External morphological traits such as genital markings have been widely applied for the identification of individual earthworm species (Briones et al. 2009). The application of such traits enables a greater degree of accuracy in species identification than alternative traits such as body size and pigmentation that can often vary within species (Piearce et al. 2002, Chang et al. 2007). The new morphological trait uncovered in the present study could therefore be a suitable morphological trait for identifying the two lineages.

One possible explanation for the existence of cryptic lineages within British earthworm species is that they may have diverged in allopatry after inhabiting separate glacial refugia during the Pleistocene (King et al. 2008). Many mtDNA-based phylogeographic studies of north European animal and plant populations have uncovered evidence for genetic lineages that are believed to have resulted from the genetic divergence of separate southern European refugial populations during the Pleistocene (Taberlet et al. 1998, Hewitt 2001). However these lineages are typically allopatric in their distribution reflecting their different routes of recolonisation. It is possible however that human-mediated dispersal may have mixed geographical lineages of *L. rubellus* given the extent of the contemporary human-mediated dispersal of earthworms and the many peregrine populations of *L. rubellus* (Schwert, 1980, Tiunov et al., 2006, Cameron et al., 2008). However, it has been proposed that the extent of such mixing would be insufficient to cause the widespread co-occurrence of lineages that is observed within many European earthworm species (King et al., 2008).

The recent discovery of genetic lineages within north European species that appear to be too highly differentiated and localised in distribution to have originated from recolonising Southern European refugial populations has led some researchers to speculate that they may have survived the Pleistocene glaciations within 'cryptic' northern refugia (Vialatte et al. 2008, Michl et al. 2010). Such refugia could have consisted of sheltered valleys with microclimates favourable to temperate adapted species, ice-free nunataks atop mountains or cave environments (Pielou 1991, Stewart and Lister 2001). Populations inhabiting such refugia may have genetically diverged from other populations due to the occurrence of a genetic bottleneck or genetic adaptation to a new ecological niche that in some circumstances may have resulted in allopatric speciation (Stewart et al. 2010). The discovery of populations of *L. rubellus* within cave environments featuring none of the woody organic matter typically associated with epigeic habitats (Reeves 1999) may indicate that this species could have adjusted its ecological niche and inhabited similar environments as cryptic refugia.

However, a refugial population of *L. rubellus* could theoretically have existed within southern England during the Pleistocene. Geological studies indicate that a large part of the south of England was unglaciated during this time-period (Sejrup et al. 2005). Millennial fluctuations in

climate due to Dansgaard-Oeschger cycles may have also resulted in a climate that was sporadically temperate in nature (Scourse et al. 2009). The existence of a land-bridge to mainland Europe may have also enabled *L. rubellus* individuals to migrate to warmer regions during the colder periods of Dansgaard-Oeschger cycles. Alternatively individuals may have survived the colder periods within refugia featuring favourable microclimates. Studies have uncovered evidence that such refugia may have existed within sheltered areas of southwest England (Stewart and Lister 2001).

A microsatellite analysis of (lineage A) *L. rubellus* populations located along a gradient of nickel contamination did not uncover any strong support for the hypothesis that the most heavily contaminated sites consist of genetically adapted populations. The analysis was unable to uncover a clear correlation between the genetic differentiation of the sites and their degree of nickel contamination. Although a slightly higher degree of genetic differentiation was uncovered for two of the contaminated sites this differentiation was not particularly pronounced when it was compared with that between different regions. This contrasted with previous studies of *L. rubellus* populations inhabiting heavy metal contaminated sites that have uncovered evidence of genetically differentiated populations (Peles et al. 2003, Kille et al. 2010, Simonsen et al. 2010). Although it is possible that the degree of resolution offered by microsatellite analysis may have been insufficient to resolve patterns of weak selection affecting only a small number of loci, a low degree of differentiation was also uncovered in a multilocus AFLP analysis of the same populations (Gabriela Juma pers. comm.). It is therefore possible that the presence of *L. rubellus* individuals at nickel-contaminated sites may be due to factors other than the genetic adaptation of populations.

The lack of a clear correlation between nickel contamination and population differentiation also did not offer strong support for the alternative hypothesis that the highly contaminated sites may be 'sink' populations sustained by the immigration of non-resistant individuals from less contaminated sites. If the highly contaminated sites contained an increased proportion of immigrant individuals, then the average genetic differentiation of these sites would have been expected to be lower than the less contaminated sites. The levels of genetic diversity at the most heavily contaminated sites also did not appear to be higher than the other sites, as they might be if these sites contained a large proportion of immigrant individuals.

Several other population genetic studies that have applied microsatellite analysis to infer the genetic adaptation of phenotypically distinct populations have also failed to uncover clear evidence of an increased genetic differentiation of the populations from surrounding populations (Jordan et al. 2005, Whiteley et al. 2009). This has led researchers to propose that these populations may have adapted through a process of phenotypic plasticity.

The phenotypic adaptation of *L. rubellus* individuals to contaminants by physiological acclimation is well documented within ecotoxicological studies (reviewed in Brulle et al. 2010). Such studies have long demonstrated that metal exposed earthworms may display physiological changes following a period of exposure to toxicants. Physiological changes form the basis of many of the biomarkers that are applied to indicate stress responses within contaminant exposed earthworms (Scott-Fordsmand and Weeks 2000). Although some of these physiological changes reflect the cellular damage caused by toxicants, some of the changes have been demonstrated to represent the activities of tolerance mechanisms. Such changes include the increased expression of metal-sequestering proteins such as metallothionein (Spurgeon et al. 2004, Asensio et al. 2007) and detoxifying enzymes such as glutathione S-transferase (LaCourse et al. 2009). The increasing application of 'omics' technologies within ecotoxicology has also enabled the resolution of complex detoxification pathways within *L. rubellus* by enabling the identification of the large numbers of transcriptional and metabolic changes occurring within toxicant-exposed individuals (Bundy et al. 2008, Owen et al. 2008).

Implications of the findings

The finding that the two lineages of *L. rubellus* are genetically differentiated in both their mitochondrial and nuclear genomes may indicate that they are reproductively isolated (although see Andre et al. 2010) and should therefore be treated as if they are separate species in accordance with the Biological Species Concept. Although some taxonomists oppose the adoption of such a genetic approach to the definition of earthworm species (Blakemore et al. 2010), treating the lineages as separate species could serve an important practical role in preventing inaccuracies arising within ecotoxicological studies. Studies of other cryptic annelid species including the species complexes of *Tubifex tubifex* (Sturmbauer et al. 1999) and *Capitella capitata* (Linke-Gamenick et al. 2000) have demonstrated that cryptic species can differ considerably in their tolerance of contaminants. It is therefore possible that the two lineages of *L. rubellus* may differ in tolerance towards some contaminants with consequences for the application of *L. rubellus* as a sentinel species. Further ecotoxicological studies are therefore required to analyse any possible differences in tolerance between lineages.

If the two lineages of *L. rubellus* are found to differ significantly in their tolerance of contaminants, then it will be necessary to account for this difference during ecotoxicological testing. Future studies featuring *L. rubellus* will therefore need to differentiate the two lineages or apply a single lineage within their studies. It is possible that one of the lineages may be better suited to ecotoxicological testing by displaying a greater sensitivity to contaminants or displaying a lower variability in response. Differentiation of the lineages could be conducted using either the

morphological traits uncovered in this study or the restrictment length polymorphism found to differentiate the lineages (Andy King personal communication).

One implication of the lack of genetic differentiation indicated by the microsatellite analysis is that *L. rubellus* may display a much greater capacity for tolerating toxic metals through phenotypic plasticity than previous studies have indicated. One way by which a species could acclimate towards the presence of contaminants could be by a process of developmental plasticity with irreversible changes during development resulting in adult individuals being more tolerant of contaminants (Meyers and Bull 2002). Existing exposure trials featuring *L. rubellus* may have failed to take this source of phenotypic plasticity into account by focusing upon the responses of adult individuals in exposure trials (e.g. Lister et al. 2011). It is therefore possible that the actual degree of phenotypic plasticity demonstrated by *L. rubellus* populations that are exposed to contaminants could be greater than is currently indicated.

Future directions

As has been mentioned the cryptic genetic lineages within *L. rubellus* could pose a severe problem for ecotoxicologists. It is therefore important that further studies of the two lineages are conducted to enable them to be understood in greater detail. One such study could be to conduct breeding trials to establish the source of the reproductive isolation observed between the two lineages, as have been performed with different lineages of *Allolobophora chlorotica* (Lowe and Butt 2008). This would also allow researchers to establish if a low level of hybridisation between the lineages does exist as suggested by the study of Andre et al. (2010). Studies could also be conducted to uncover differences in feeding ecology between the lineages that could indicate differences in their ecological niche (Curry and Schmidt 2007). Such differences in feeding could lead to differences in their degree of exposure towards certain contaminants. Finally in order to establish how widely the lineages of *L. rubellus* differ in their tolerance towards contaminants, comparative exposure trials are required that could indicate differences in the uptake, excretory ability or assimilation of contaminants.

Further analysis is also required to determine how the *L. rubellus* populations at Clydach are able to tolerate the toxic levels of nickel found there. As has been mentioned, the design of existing nickel exposure trials featuring *L. rubellus* (Lister et al. 2011) have failed to take into account the possible effects of developmental plasticity in increasing tolerance. An exposure trial could therefore be conducted featuring *L. rubellus* cocoons or juveniles to investigate this possibility.

The possible absence of a genetically adapted population at Clydach could also be further investigated to establish the circumstances that may have prevented its evolution. Comparisons could be made with populations at other nickel contaminated sites with population studies

conducted to establish if any of these sites feature genetically adapted populations. Factors that could influence adaptation and therefore could be taken into account in selecting the sites include the level of genetic diversity in nearby populations, the duration of the nickel contamination and the presence of additional physical stressors. If a population is found to feature genetically adapted individuals a study could also be conducted to determine if there are any 'costs-of-tolerance' associated with this genetic tolerance that could constrain its development within other populations (Posthuma and Van Straalen 1993).

Finally given that the majority of the microsatellites applied in the present study amplify both lineages of *L. rubellus*, a population genetics study could be undertaken to determine the effects of contaminants upon populations of both lineages. This would enable a comparison of the effects upon the two lineages and could uncover differences between them, e.g. adaptation in only a single lineage.

Bibliography

Adamowicz, S.J., Menu-Marque, S., Hebert, P.D.N. and Purvis, A. (2007) 'Molecular systematics and patterns of morphological evolution in the *Centropagidae* (Copepoda: Calanoida) of Argentina', *Biological Journal of the Linnean Society*, 90, pp. 279–292.

Addison, J.A. (2009) 'Distribution and impacts of invasive earthworms in Canadian forest ecosystems', *Biological Invasions* 11, pp. 59-79.

Al-Hiyaly, S.A.K., McNeilly, T., Bradshaw, A.D. (1988) 'The effect of zinc contamination from electricity pylons – evolution in a replicated situation', *New Phytologist* 110, pp. 571–580.

Al-Hiyaly, S.A.K., McNeilly, T., Bradshaw, A.D. and Mortimer, A.M. (1993) 'The effect of zinc contamination from electricity pylons. Genetic constraints on selection for zinc tolerance', *Heredity* 70, pp. 22–32.

Anderson, E.C. and Thompson, E.A. (2002) 'A model-based method for identifying species hybrids using multilocus genetic data', *Genetics* 160, pp. 1217-1229.

Anderson, J.T., and Geber. M.A. (2009) 'Demographic source-sink dynamics restrict local adaptation in Elliott's blueberry (*Vaccinium elliottii*)', *Evolution* 64, pp. 370-384.

Anderson, S., Sadinski, W., Shugart, L., Brussard, P., Depledge, M., Ford, T., Hose, J., Stegeman, J., Suk, W., Wirgin, I. and Wogan, G. (1994) 'Genetic and molecular ecotoxicology: a research framework', *Environmental Health Perspectives*, 102 (Suppl 12), pp. 3–8.

Andre, J., King, R. A., Sturzenbaum, S. R., Kille, P., Hodson, M. E. and Morgan, A. J. (2010) 'Molecular genetic differentiation in earthworms inhabiting a heterogeneous Pb-polluted landscape', *Environmental Pollution*, 158, pp. 883-890.

Arnold, B.E., Hodson, M.E., Charnock, J. and Peijnenburg, W. (2008) 'Comparison of subcellular partitioning, distribution, and internal speciation of Cu between Cu-tolerant and Naive populations of *Dendrodrilus rubidus* Savigny', *Environmental Science and Technology* 42, pp. 3900–3905.

Asensio, V., Kille, P., Morgan, A.J., Soto, M. and Marigomez, I. (2007) 'Metallothionein expression and Neutral Red uptake as biomarkers of metal exposure and effect in *Eisenia fetida* and *Lumbricus rubellus* exposed to Cd', *European Journal of Soil Biology* 43, S233-S238.

Avise, J.C. (2000) *Phylogeography*. Cambridge, MA: Harvard University Press.

Avise, J.C. and Walker, D. (1999) 'Species realities and numbers in sexual vertebrates: perspectives from an asexually transmitted genome', *Proceedings of the National Academy of Sciences of the United States of America* 96, pp. 992–995.

Babczyńska, A., Wilczek, G. and Migula, P. (2010) 'Effects of dimethoate on spiders from metal pollution gradient', *The Science of the Total Environment* 370, pp. 352-359.

Bach, L., Palmqvist, A., Rasmussen, L.J. and Forbes, V.E. (2005) 'Differences in PAH tolerance between *Capitella* species: Underlying biochemical mechanisms', *Aquatic Toxicology*, 74, pp. 307-319.

Baker, A.M., Bartlett, C., Bunn, S.E., Goudkamp, K., Sheldon, F. and Hughes, J.M. (2003) 'Cryptic species and morphological plasticity in long-lived bivalves (Unionoida: Hyriidae) from inland Australia', *Molecular Ecology* 12, pp. 2707–2717.

Ballard, J.W.O. and Whitlock, M.C. (2004) 'The incomplete natural history of mitochondria', *Molecular Ecology*, 13, pp. 729 – 744.

Beaumont, M.A. and Bruford, M.W. (1999) 'Microsatellites in Conservation Genetics', in Goldstein, D.B. and Schlötterer, C. (eds.) *Microsatellites: Evolution and applications*. New York: Oxford University Press, pp. 165-182.

Beeby, A.N. (1991) 'Toxic metal uptake and essential metal regulation in terrestrial invertebrates: a review', in Newman, M.C. and McIntosh, A.W. (eds.) *Metal Ecotoxicology: Concepts and Applications*. Chelsea, MI: Lewis, pp. 65–90.

Beldade, P., Mateus, A.R.A. and Keller, R.A. (2011) 'Evolution and molecular mechanisms of adaptive developmental plasticity', *Molecular Ecology* 20, pp. 1347-1363.

Belfiore, N.M. and Anderson, S.L. (2001) 'Effects of contaminants on genetic patterns in aquatic organisms: a review', *Mutation Research* 489, pp. 97–122.

Belkhir, K., Borsa, P., Chikhi, L., Raufaste, N. and Bonhomme, F. (2004). GENETIX 4.05, Logiciel Sous Windows TM Pour la Génétique des Populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier.

Bely, A.E. and Wray, G.A. (2004) 'Molecular phylogeny of naidid worms (Annelida; Clitellata) based on cytochrome c oxidase I', *Molecular Phylogenetics and Evolution* 30, pp. 50–63.

Berckmoes, V., Scheirs, J., Jordaens, K., Blust, R., Backeljau, T. and Verhagen, R. (2005) 'Effects of environmental pollution on microsatellite DNA diversity in wood mouse (*Apodemus sylvaticus*) populations', *Environmental Toxicology and Chemistry* 24, 2898–2907.

Bickford, D., Lohman, D. J., Sodhi, N. S., Ng, P. K. L., Meier, R., Winker, K., Ingram, K. K. and Das, I. (2007) 'Cryptic species as a window on diversity and conservation', *Trends in Ecology and Evolution*, 22, pp. 148-155.

Bickham, J.W., Sandhu, S., Hebert, P.D.N., Chikhi, L. and Athwal, R. (2000) 'Effects of chemical contaminants an genetic diversity in natural populations: implications for biomonitoring and ecotoxicology', *Mutation Research* 463, pp. 33–51.

Binns, J. (2004) 'Remediation: Cleanup complete at Love Canal', *Civil Engineering-ASCE*, 74, pp. 22-23.

Blakemore, R.J., Kupriyanova, E.K. and Grygier, M.J. (2010) 'Neotypification of *Drawida hattamimizu* Hatai, 1930 (Annelida, Oligochaeta, Megadrili, Moniligastridae) as a model linking mtDNA (COI) sequences to an earthworm type, with a response to the 'Can of Worms' theory of cryptic species', *ZooKeys* 41, pp. 1–29.

Blanquer-Maumont, A. and Crouau-Roy B. (1995) 'Polymorphism, monomorphism, and sequences in conserved microsatellites in primate species', *Journal of Molecular Ecology* 41, pp. 492–497.

Blouin, M., Zuily-Fodil, Y., Pham-Thi, A.T., Laffray, D., Reversat, G., Pando, A., Tondoh, J. and Lavelle, P. (2005) 'Belowground organism activities affect plant aboveground phenotype, inducing plant tolerance to parasites', *Ecology Letters* 8, pp. 202-208.

Bolnick, D.I. (2004) 'Does intraspecific competition generate disruptive selection? An experimental test in natural populations of sticklebacks', *Evolution* 58, pp. 608-618.

Booth, L.H. and O'Halloran, K. (2001) 'A comparison of biomarker responses in the earthworm Aporrectodea caliginosa to the organophosphorous insecticides Diazinon and Chlorpyrifos', *Environmental Toxicology and Chemistry* 20, pp. 2494–2502.

Bouché, M.B. (1972) 'Lombriciens de France. Ecologie et Systématique', *INRA Annales de Zoologie, Ecologie Animale*, 72, pp. 1-671.

Bouché, M.B. (1977) 'Strategies Lombriciennes', *Ecological Bulletin* 25, pp. 122-132.

Bouché, M. B. (1992) 'Earthworm species and ecotoxicological studies', in Greig-Smith, P.W., Becker, H., Edwards, P.J. and Heimbach, F. (eds.) *Ecotoxicology of Earthworms*. Andover: Intercept Ltd., pp. 20-35.

Bourret, V., Couture, P., Campbell, P.G.C. and Bernatchez, L. (2008) 'Evolutionary ecotoxicology of wild yellow perch (*Perca flavescens*) populations chronically exposed to a poly- metallic gradient', *Aquatic Toxicology* 86, pp. 76–90.

Briones, M.J.I., Morán, P. and Posada, D. (2009) 'Are the sexual, somatic and genetic characters enough to solve nomenclatural problems in lumbricid taxonomy?', *Soil Biology and Biochemistry* 41, pp. 2257–2271.

Brohede, J. and Ellegren, H. (1999) 'Microsatellite evolution: polarity of substitutions within repeats and neutrality of flanking sequences', *Proceedings of the Royal Society B: Biological Sciences* 266, pp. 825–833.

Brown, G.G. and Doube, B.M. (2004) 'Functional interactions between earthworms, microorganisms, organic matter, and plants', in Edwards, C.A. (ed.) *Earthworm Ecology*. Boca Raton, FL: CRC Press, pp. 213–224.

Brulle, F., Morgan, A.J., Cocquerelle, C. and Vandenbulcke, F. (2010) 'Transcriptomic underpinning of toxicant-mediated physiological function alterations in three terrestrial invertebrate taxa: a review', *Environmental Pollution* 158, pp. 2793-3014.

Bundy, J.G., Keun, H., Sidhu, J.K., Spurgeon, D.J., Svendsen, C., Kille, P. and Morgan, A.J. (2007) 'Metabolic profile biomarkers of metal contamination in a sentinel terrestrial species are applicable across multiple sites', *Environmental Science and Technology* 41, pp. 4458-4464.

Bundy, J.G., Sidhu, J.K., Rana, F., Spurgeon, D.J., Svendsen, C., Wren, J., Stürzenbaum, S.R., Morgan, A.J. and Kille, P. (2008) ''Systems toxicology' approach identifies coordinated metabolic responses to copper in a terrestrial non-model invertebrate, the earthworm *Lumbricus rubellus*', *BMC Biology* 6, 25.

Bundy, J.G., Spurgeon, D.J., Svendsen, C., Hankard, P. K., Weeks, J. M., Osborn, D., Lindon, J.C. and Nicholson, J.K. (2004) 'Environmental metabonomics: applying combination biomarker analysis in earthworms at a metal contaminated site', *Ecotoxicology* 13, pp. 797–806.

Burbrink, F.T. and Pyron, R.A. (2010) 'How does ecological opportunity influence rates of speciation, extinction, and morphological diversification in New World ratsnakes (tribe Lampropeltini)?', *Evolution* 64, pp. 934-943.

Burke, J.M. and Arnold, M.L. (2001) 'Genetics and the fitness of hybrids', *Annual Review of Genetics* 35, pp. 31-52.

Butler, G. C. (1978) *Principles of Ecotoxicology*. New York: Wiley & Sons.

Calabrese, P.P. and Sainudiin, R. (2004) 'Models of microsatellite evolution', in Nielsen, R. (ed.) *Statistical Methods in Molecular Evolution, Series: Statistics for Biology and Health*. New York: Springer, pp. 289-306.

Cameron, E.K., Bayne, E.M. and Coltman, D.W. (2008) 'Genetic structure of invasive earthworms *Dendrobaena octaedra* in the boreal forest of Alberta: insights into introduction mechanisms', *Molecular Ecology* 17, pp. 1189-1197.

Campana, C., Gauvin, S. and Ponge, J.-F. (2002) 'Influence of ground cover on earthworm communities in an unmanaged beech forest: linear gradient studies', *European Journal of Soil Biology* 38, pp. 213-224.

Cannavacciuolo, M., Bellido, A., Cluzeau, D., Gascuel, C. and Trehen, P. (1998) 'A geo- statistical approach to the study of earthworm distribution in grassland', *Applied Soil Ecology* 9, pp. 345-349.

Caro, T.M. and O'Doherty, G. (1999) 'On the use of surrogate species in conservation biology', *Conservation Biology* 13, pp. 805-814.

Carpenè, E., Andreani, G., Monari, M., Castellani, G. and Isani, G. (2006) 'Distribution of Cd, Zn, Cu and Fe among selected tissues of the earthworm (*Allolobophora caliginosa*) and Eurasian woodcock (*Scolopax rusticola*)', *Science of the Total Environment* 363, pp. 126-35.

Carson, R.L. (1962) *Silent spring*. Boston, M.A: Houghton Mifflin.

Chang, C.-H., Lin, Y.-H., Chen, I.-H., Chuang, S.-C. and Chen, J.-H. (2007) 'Taxonomic re-evaluation of the Taiwanese montane earthworm *Amynthas wulinensis* Tsai, Shen & Tsai, 2001 (Oligochaeta: Megascolecidae): Polytypic species or species complex?', *Organisms Diversity & Evolution* 7, pp. 231.

Chang, C.-H., Lin S.-M., and Chen J.-H. (2008) 'Molecular systematics and phylogeography of the gigantic earthworms of the *Metaphire formosae* species group (Clitellata, Megascolecidae)', *Molecular Phylogenetics and Evolution* 49, pp. 958-968.

Chang, C. H., Rougerie, R. and Chen, J.H. (2009) 'Identifying earthworms through DNA barcodes: pitfalls and promise' *Pedobiologia* 52, pp. 171-180.

Chapman, P. M. (1995) 'Ecotoxicology and pollution – key issues', *Marine Pollution Bulletin* 31, pp. 4–12.

Chapman, P.M. (2002) 'Integrating toxicology and ecology: putting the "eco" into ecotoxicology', *Marine Pollution Bulletin* 44, pp. 7-15.

Chenon, P., Rousset, A. and Crouau, Y., (2000) 'Genetic polymorphism in nine clones of a parthenogenetic collembolan used in ecotoxicological testing', *Applied Soil Ecology* 14, pp. 103–110.

Chistiakov, D.A., Hellemans, B., Filip, A.M. and Volckaert, F.A.M. (2006) 'Microsatellites and their genomic distribution, evolution, function and applications: A review with special reference to fish genetics', *Aquaculture* 255, pp. 1–29.

Clapp, B.W. (1994) *An environmental history of Britain since the industrial revolution.* London: Longman.

Coe, T.S., Hamilton, P.B., Griffiths, A.M., Hodgson, D.J., Wahab, M.A. and Tyler, C.R. (2009) 'Genetic variation in strains of zebrafish (*Danio rerio*) and the implications for ecotoxicology studies', *Ecotoxicology* 18, pp. 144-150.

Crispo, E. (2008) 'Modifying effects of phenotypic plasticity on interactions among natural selection, adaptation and gene flow', *Journal of Evolutionary Biology* 21, pp. 1460–1469.

Curry, J.P. and Schmidt, O. (2007) 'The feeding ecology of earthworms- a review', *Pedobiologia* 50, pp. 463-477.

Dai, J., Becquer, T., Henri Rouiller, J., Reversat, G., Bernhard-Reversat, F., Nahmani, J. and Lavelle, P. (2004) 'Heavy metal accumulation by two earth- worm species and its relationship to total and DTPA-extractable metals in soils', *Soil Biology and Biochemistry* 36, pp. 91-98.

Damgaard, J., Klass, K-D., Picker, M.D. and Buder, G., 2008. 'Phylogeny of the Heelwalkers (Insecta: Mantophasmatodea) based on mtDNA sequences, with evidence for additional taxa in South Africa', *Molecular Phylogenetics and Evolution* 47, pp. 443–462.

Davison, A. (2006) 'The ovotestis: an underdeveloped organ of evolution', *BioEssays* 28, pp. 642-50.

Depledge, M.H., Aagaard, A. and Györkös, P. (1995) 'Assessment of trace metal toxicity using molecular, physiological and behavioural biomarkers', *Marine Pollution Bulletin* 31, pp. 19–27.

Dettman, J.R., Jacobson, D.J., Turner, E., Pringle, A., and Taylor, J.W. (2003) 'Reproductive isolation and phylogenetic divergence in *Neurospora*: comparing methods of species recognition in a model eukaryote', *Evolution* 57, pp. 2721–2741.

Dettman, J.R. and Taylor, J.W. (2004) 'Mutation and evolution of microsatellite loci in Neurospora', *Genetics* 168, pp. 1231– 1248.

Dias, P. C. (1996) 'Sources and sinks in population biology', *Trends in Ecology and Evolution* 11, pp. 326–330.

Dimitrov, D., Arnedo, M.A. and Ribera C. (2008) 'Colonization and diversification of the spider genus *Pholcus* Walckenaer, 1805 (Araneae, Pholcidae) in the Macaronesian archipelagoes: Evidence for long-term occupancy yet rapid recent speciation', *Molecular Phylogenetics and Evolution* 42, pp. 596-614.

Diogo, J.B., Natal-da-Luz, T., Sousa, J.P., Vogt, C. and Nowak, C. (2007) 'Tolerance of genetically characterized *Folsomia candida* strains to phenmedipham exposure a comparison between reproduction and avoidance tests', *Journal of Soils and Sediments* 7, pp. 388-392.

Domingo-Roura, X., Lopez-Giraldez, F., Saeki, M. and Marmi, J. (2005) 'Phylogenetic inference and comparative evolution of a complex microsatellite and its flanking regions in carnivores', *Genetical Research Cambridge* 85, pp. 223-233.

Domínguez, J., Velando, A. and Ferreiro, A. (2005) 'Are Eisenia fetida (Savigny, 1826) and Eisenia andrei Bouché (1972) (Oligochaeta, Lumbricidae) different biological species?', *Pedobiologia* 49, pp. 81–87.

Dupont, L., Lazrek, F., Porco, D., King, R.A., Rougerie, R., Symondson, W.O.C., Livet, A., Richard, B., Decaëns, T., Butt, K.R. and Mathieu, J. (2011) 'New insight into the genetic structure of the *Allolobophora chlorotica* aggregate in Europe using microsatellite and mitochondrial data', *Pedobiologia* 54, pp. 217-224.

Eason, C. and O'Halloran, K. (2002) 'Biomarkers in toxicology versus ecological risk assessment', *Toxicology* 182, pp. 517-521.

Edwards, C.A. and Bohlen, P.J. (1992) 'The effects of toxic chemicals on earthworms', *Reviews of Environmental Contamination and Toxicology* 125, pp. 23–100

Edwards, C.A. and Bohlen, P.J. (1996) *Biology and ecology of earthworms*. London: Chapman & Hall.

Eijsackers, H. (2010) 'Earthworms as colonisers: primary colonisation of contaminated land, and sediment and soil waste deposits', *The Science of the Total Environment* 408, pp. 1759-1769.

Eisen, J.A., (1999) 'Mechanistic basis for microsatellite instability', in Goldstein, D.B. and Schlötterer, C. (eds.) *Microsatellites: evolution and applications*. New York: Oxford University Press, pp. 34-48.

Elmer, K.R., Dávila, J.A. and Lougheed, S.C. (2007) 'Cryptic diversity and deep divergence in an upper Amazonian leaflitter frog, *Eleutherodactylus ockendeni*', *BMC Evolutionary Biology* 7, 247.

Elmer, K.R., Lehtonen, T.K., Kautt, A.F., Harrod, C. and Meyer, A. (2010) 'Rapid sympatric ecological differentiation of crater lake cichlid fishes within historic times', *BMC Biology* 8, 60.

Elvira, C., Domínguez, J. and Mato, S. (1996) 'The growth and reproduction of *Lumbricus rubellus* and *Dendrobaena rubida* in cow manure Mixed cultures with *Eisenia andrei*', *Applied Soil Ecology* 5, pp. 97-103.

Enckell, P.H., Niklasson, M., Stille, B. and Douwes, P. (1986) 'Insulation and isolation factors influencing the genetic variation in *Lumbricus rubellus* Hoffm. (Lumbricidae) in the Faroe Islands', *Hereditas* 104, pp. 263–272.

Environment Agency (2009), *Soil guideline values for nickel in soil.* Bristol: Environment Agency.

Eränen, J.K., Nilsen, J., Zverev, V.E. and Kozlov, M.V. (2009) 'Mountain birch under multiple stressors- heavy metal-resistant populations co-resistant to biotic stress but maladapted to abiotic stress', *Journal of Evolutionary Biology* 22, pp. 840-851.

Erséus, C. and Gustafsson, D. (2009) 'Cryptic speciation in clitellate model organisms', in Shain, D. (ed.) *Annelids in modern biology.* Hoboken, NJ: Wiley-Blackwell, pp. 31-46.

Estoup, A., Tailliez, C., Cornuet, J.M., and Solignac, M. (1995) 'Size homoplasy and mutational processes of interrupted microsatellites in two bee species, *Apis mellifera* and *Bombus terrestris* (Apidae)' *Molecular Biology and Evolution* 12, pp. 1074-1084.

Evanno, G., Regnaut, S., and Goudet, J. (2005) 'Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study', *Molecular Ecology* 14, pp. 2611-2620.

Evenden, A.J. and Depledge, M.H. (1997) 'Genetic susceptibility in ecosystems: the challenge for ecotoxicology', *Environmental Health Perspectives* 105 Suppl 4, pp. 849-854.

Excoffier, L., Laval, G. and Schneider, S. (2005) 'ARLEQUIN (version 3.0): An integrated software package for population genetics data analysis', *Evolutionary Bioinformatics Online* 1, pp. 47–50.

Fazio, V.W., Miles, D.B. and White. M.W. (2004) 'Genetic differentiation in the endangered black-capped vireo', *Condor* 106, pp. 377-385.

Fedorenkova, A., Vonk, J.A., Lenders, H.J.R., Ouborg, N.J., Breure, A.M. and Hendriks, A.J. (2010) 'Ecotoxicogenomics: bridging the gap between genes and populations', *Environmental Science and Technology* 44, pp. 4328–4333.

Feulner, P.G.D., Kirschbaum, F., Schugardt, C., Ketmaier, V. and Tiedemann, R. (2006) 'Electrophysiological and molecular genetic evidence for sympatrically occuring cryptic species in African weakly electric fishes (Teleostei: Mormyridae: Campylomormyrus)', *Molecular Phylogenetics and Evolution* 39, pp. 198–208.

Fisher, M.C., Koenig, G.L., White, T.J., and Taylor, J.W. (2000) 'A test for concordance between the multilocus genealogies of genes and microsatellites in the pathogenic fungus *Coccidioides immitis*', *Molecular Biology and Evolution* 17, pp. 1164–1174.

Fisker, K.V., Sørensen, J.G., Damgaard, C., Ladegaard Pedersen, K., and Holmstrup, M. (2011) 'Genetic adaptation of earthworms to copper pollution: is adaptation associated with fitness costs in *Dendrobaena octaedra*?', *Ecotoxicology* 20, pp. 563-573.

Folmer, O., Black, M., Hoeh, W., Lutz, R. and Vrijenhoek, R. (1994) 'DNA primers for amplification of mitochondrial cytochrome *c* oxidase subunit I from diverse metazoan invertebrates', *Molecular Marine Biology and Biotechnology* 3, pp. 294-299.

Forbes, V.E. (1998) 'Sources and implications of variability in sensitivity to chemicals for ecotoxicological risk assessment', *Archives of Toxicology Supplement* 20, pp. 407-418.

Forbes, V.E., Palmqvist, A. and Bach, L. (2006) 'The use and misuse of biomarkers in ecotoxicology', *Environmental Toxicology & Chemistry* 25, pp. 272-280.

Francioni, E. and Wagener, A.D.L.R. (2007) 'Evaluation of the mussel *Perna perna* as a biomonitor of polycyclic aromatic hydrocarbon (PAH) exposure and effects', *Marine Pollution Bulletin* 54, pp. 329-338.

Fränzle, O. (2006) 'Complex bioindication and environmental stress assessment', *Ecological Indicators* 6, pp. 114–136.

Frati, F., Negri, I., Fanciulli, P.P., Pellecchia, M., De-Paola, V., Scali, V. and Dallai, R. (2004) 'High levels of genetic differentiation between Wolbachia-infected and non-infected populations of Folsomia candida (Collembola, Isotomidae)', *Pedobiologia* 48, pp. 461-468.

Galay-Burgos, M., Spurgeon, D.J., Weeks, J.M., Sturzenbaum, S.R., Morgan, A.J. and Kille, P. (2003) 'Developing a new method for soil pollution monitoring using molecular genetic biomarkers', *Biomarkers* 8, pp. 229–239.

Garza, J.C., Slatkin, M. and Freimer, N.B. (1995) 'Microsatellite allele frequencies in humans and chimpanzees, with implications for constraints on allele size', *Molecular Biology and Evolution* 12, pp. 594–603.

Glaubitz, J.C. (2004) 'CONVERT: a user friendly program to reformat diploid genotypic data for commonly used population genetic software packages', *Molecular Ecology Notes* 4, pp. 309-310.

Gonzalez-Rey, M., Serafim, A., Company, R. and Bebianno, M.J. (2007) 'Adaptation to metal toxicity: a comparison of hydrothermal vent and coastal shrimps', *Marine Ecology* 28, pp. 100-107.

Goto, D. and Wallace W.G. (2010) 'Metal intracellular partitioning as a detoxification mechanism for mummichogs (*Fundulus heteroclitus*) living in metal-polluted salt marshes', *Marine Environmental Research* 69, pp. 163-171.

Grassle, J.P. and Grassle, J.F. (1976) 'Sibling species in the marine pollution indicator, *Capitella capitata* (Polychaeta)', *Science* 192, pp. 567-569.

Great Britain. House of Commons Select Committee on the Environment (1990) *Contaminated Land First. Report, Session 1989-90, HC170.* London: HMSO.

Grimsrud, T.K. and Peto, J. (2006) 'Persisting risk of nickel related lung cancer and nasal cancer among Clydach refiners', *Occupational and Environmental Medicine* 63, pp. 365-366.

Guana, F., Niua, A-O, Attwood, S.W., Li, Y.-L., Zhang, B. and Zhu, Y.-H. (2008) 'Molecular phylogenetics of Triculine snails (Gastropoda: Pomatiopsidae) from southern China', *Molecular Phylogenetics and Evolution* 48, pp. 702–707.

Guindon, S. and Gascuel O. (2003) 'A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood', *Systematic Biology* 52, pp. 696-704.

Hajibabaei, M., Singer, G.A.C., Hebert, P.D.N. and Hickey, D. (2007) 'DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics', *Trends in Genetics*, 23, pp. 167-172.

Hall, T.A. (1999) 'BIOEDIT: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT', *Nucleic Acids Symposium Series* 41, pp. 95–98.

Hamers, T., Van den Berg, J.H.J., Van Gestel, C.A.M., Van Schooten, F.-J. and Murk, A.J. (2006) 'Risk assessment of metals and organic pollutants for herbivorous and carnivorous small mammal food chains in a polluted floodplain (Biesbosch, The Netherlands)', *Environmental Pollution* 144, pp. 581–595.

Hancock, J. M. (1999) 'Microsatellites and other simple sequences: genomic context and mutational mechanism' in Goldstein, D.B. and Schlötterer, C. (eds.) *Microsatellites: evolution and applications*. New York: Oxford University Press, pp. 1–9.

Hansen, P.L., Holmstrup, M., Bayley, M. and Simonsen, V. (2006) 'Low genetic variation for *Dendrobaena octaedra* from Greenland compared to populations from Europe and North America: Refuge or selection?', *Pedobiologia* 50, pp. 225-234.

Harper, G.L., Cesarini, S., Casey, S.P., Morgan, A.J., Kille, P. and Bruford, M.W. (2006) 'Microsatellite markers for the earthworm *Lumbricus rubellus*', *Molecular Ecology Notes* 6, pp. 325–327.

Hebert, P.D.N. and Luiker, M.M. (1996) 'Genetic effects of contaminant exposure—towards an assessment of impacts on animal populations', *Science of the Total Environment* 191, pp. 23–58.

Hebert, P.D.N., Penton, E.H., Burns, J.M., Janzen, D.H. and Hallwachs, W. (2004) 'Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*', *Proceedings of the National Academy of Sciences of the United States of America* 101, pp. 14812–14817.

Hebert, P.D.N., Ratnasingham, S. and deWaard, J.R. (2003) 'Barcoding animal life: cytochrome *c* oxidase subunit 1 divergences among closely related species', *Proceedings of the Royal Society of London Series B-Biological Sciences* 270, pp. S96-S99.

Heethoff, M., Domes, K., Laumann, M., Maraun, M., Norton, R.A. and Scheu, S. (2006) 'High genetic divergences indicate ancient separation of parthenogenetic lineages of the oribatid mite *Platynothrus peltifer* (Acari, Oribatida)', *Journal of Evolutionary Biology* 20, pp. 392-402.

Heethoff, M., Etzold, K. and Scheu, S. (2004) 'Mitochondrial COII sequences indicate that the parthenogenetic earthworm *Octolasion tyrtaeum* (Savigny 1826) constitutes of two lineages differing in body size and genotype', *Pedobiologia* 48, pp. 9-13.

Hendry, A.P., Taylor, E.B. and McPhail, J.D. (2002) 'Adaptive divergence and the balance between selection and gene flow: lake and stream stickleback in the Misty system', *Evolution* 56, pp. 1199-1216.

Hewitt, G.M. (2001) 'Speciation, hybrid zones and phylogeography - or seeing genes in space and time', *Molecular Ecology* 10, pp. 537-549.

Hilty, J. and Merenlender, A. (2000) 'Faunal indicator taxa selection for monitoring ecosystem health', *Biological Conservation* 92, pp. 185-197.

Hirth, J.R., Li, G.D., Chan, K.Y. and Cullis, B.R. (2009) 'Long-term effects of lime on earthworm abundance and biomass in an acidic soil on the south-western slopes of New South Wales, Australia', *Applied Soil Ecology* 43, pp. 106-114.

Hoffman, D.J., Rattner, B.A. and Hall, R.J. (1990) 'Wildlife toxicology', *Environmental Science and Technology* 24, pp. 276–283.

Huang, J., Xu, Q., Sun, Z.J., Tang, G.L. and Su, Z.Y. (2007) 'Identifying earthworms through DNA barcodes', *Pedobiologia* 51, pp. 301-309.

Ireland, M.P. (1983) 'Heavy metal uptake and tissue distribution in earthworms', in Satchell, J.E. (ed.), *Earthworm Ecology. From Darwin to Vermiculture.* London: Chapman and Hall, pp. 247-265.

James, S.W., Porco, D., Decaëns, T., Richard, B., Rougerie, R. and Erséus, C. (2010) 'DNA Barcoding Reveals Cryptic Diversity in *Lumbricus terrestris* L., 1758 (Clitellata): Resurrection of *L. herculeus* (Savigny, 1826)', *PLoS ONE* 5, e15629.

Jensen, J., Diao, X. and Scott-Fordsmand, J.J. (2007) 'Sub-lethal toxicity of the antiparasitic abamectin on earthworms and the application of neutral red retention time as a biomarker', *Chemosphere* 68, pp. 744–750.

Ji, J., Choi, H.J. and Ahn, I.Y. (2006) 'Evaluation of Manila clam *Ruditapes philippinarum* as a sentinel species for metal pollution monitoring in estuarine tidal flats of Korea: effects of size, sex, and spawning on baseline accumulation', *Marine Pollution Bulletin* 52, pp. 447-453.

Jiménez-ambriz, G., Petit, C., Bourrié, I., Dubois, S., Olivieri, I and Ronce, O. (2001) 'Life history variation in the heavy metal tolerant plant *Thlaspi caerulescens* growing in a network of contaminated and noncontaminated sites in southern France : role of gene flow, selection and phenotypic plasticity', *New Phytologist* 173, pp. 199-215.

Jordan, M.A., Snell, H.L., Snell, H.M. and Jordan, W.C. (2005) 'Phenotypic divergence despite high levels of gene flow in Galapagos lava lizards (*Microlophus albemarlensis*)', *Molecular Ecology* 14, pp. 859–867.

Jouquet, P., Dauber, J., Lagerlöf, J., Lavelle, P. and Lepage, M. (2006) 'Soil invertebrates as ecosystem engineers: intended and accidental effects on soil and feedback loops', *Applied Soil Ecology* 32, pp. 153–164.

Kamitani, T. and Kaneko, N. (2007) 'Species-specific heavy metal accumulation patterns of earthworms on a floodplain in Japan', *Ecotoxicology and Environmental Safety* 66, pp. 82-91.

Kammenga, J.E., Dallinger, R., Donker, M.J., Köhler, H.-R., Simonsen, V., Triebskorn, R. and Weeks, J.M. (2000) 'Biomarkers in terrestrial invertebrates: potential and limitations for ecotoxicological soil risk assessment', *Reviews of Environmental Contamination & Toxicology* 164, pp. 93–147.

Khan, F.R., Irving, J.R., Bury, N.R. and Hogstrand, C. (2011) 'Differential tolerance of two *Gammarus pulex* populations transplanted from different metallogenic regions to a polymetal gradient', *Aquatic Toxicology* 102, pp. 95-103.

King, R.A., Tibble, A.L. and Symondson, W.O.C. (2008) 'Opening a can of worms: unprecedented sympatric cryptic diversity within British lumbricid earthworms', *Molecular Ecology* 17, pp. 4684-4698.

Klerks, P.L. (2002) 'Adaptation, Ecological Impacts, and Risk Assessment: Insights from Research at Foundry Cove, Bayou Trepagnier, and Pass Fourchon', *Human and Ecological Risk Assessment* 8, pp. 971-982.

Klerks, P.L. and Levinton. J.S. (1989) 'Rapid evolution of metal resistance in a benthic oligochaete inhabiting a metal-polluted site', *The Biological Bulletin* 176, pp. 135-141.

Klerks, P.L. and Moreau, C.J. (2001) 'Heritability of resistance to individual contaminants and to contaminant mixtures in the sheepshead minnow (*Cyprinodon variegatus*)', *Environmental Toxicology and Chemistry* 20, pp. 1746-1751.

Knapen, D., De Wolf, H., Knaepkens, G., Bervoets, L., Eens, M., Blust, R. and Verheyen E. (2009) 'Historical metal pollution in natural gudgeon populations: Interferences from allozyme, microsatellite and condition factor analysis', *Aquatic Toxicology* 95, pp. 17-26.

LaCourse, E.J., Hernandez-Viade, M., Jefferies, J.R., Svendsen, C., Spurgeon, D.J., Barrett, J., Kille, P. and Brophy, P.M. (2009) 'Glutathione transferase (GST) as a candidate molecular-based biomarker or soil toxin exposure in the earthworm *Lumbricus rubellus*', *Environmental Pollution* 157, pp. 2459-2469.

Langdon, C.J., Morgan, A.J., Charnock, J.M., Semple, K.T. and Lowe, C.N. (2009) 'As-resistance in laboratory-reared F1, F2 and F3 generation offspring of the earthworm *Lumbricus rubellus* inhabiting an As-contaminated mine soil', Environmental Pollution 157, pp. 3114–3119.

Langdon, C.J., Piearce, T.G., Black, S. and Semple, K.T. (1999) 'Resistance to arsenic toxicity in a population of the earthworm *Lumbricus rubellus*', *Soil Biology and Biochemistry* 31, pp. 1963–1967.

Langdon, C.J., Piearce, T.G., Meharg, A.A. and Semple, K.T. (2001) 'Survival and behaviour of the earthworms *Lumbricus rubellus* and *Dendrodrilus rubidus* from arsenate- contaminated and non-contaminated sites', *Soil Biology and Biochemistry* 33, pp. 1239–1244.

Langdon, C.J., Piearce, T.G., Meharg, A.A. and Semple, K.T. (2003) 'Inherited resistance to arsenate toxicity in two populations of *Lumbricus rubellus*', Environmental Toxicology and Chemistry 22, pp. 2344–2348.

Lanno, R., Wells, J., Conder, J., Bradham, K. and Basta, N. (2004) 'The bioavailability of chemicals in soil for earthworms', *Ecotoxicology and Environmental Safety* 57, pp. 39-47.

Łaszczyca, P., Augustyniak, M., Babczyńska, A., Bednarska, K., Kafel, A., Migula, P., Wilczek, G. and Witas, I. (2004) 'Profiles of enzymatic activity in earthworms from zinc, lead and cadmium polluted areas near Olkusz (Poland)', *Environment International* 30, pp. 901–910.

Lavelle, P., Bignell, D.E., Lepage, M., Volters, W., Roger, P., Ineson, P., Heal, W. and Dillion, S. (1997) 'Soil function in a changing world: the role of invertebrate ecosystem engineers', *European Journal of Soil Biology* 33, pp. 159-193.

LeBlanc, G.A. and Bain, L.J. (1997) 'Chronic Toxicity of Environmental Contaminants: Sentinels and Biomarkers', *Environmental Health Perspectives* 105 (Suppl. 1), pp. 65-79.

Lefébure, T., Douady, C.J., Gouy, M., Trontelj, P., Briolay, J. and Gibert, J. (2006) 'Phylogeography of a subterranean amphipod reveals cryptic diversity and dynamic evolution in extreme environments', *Molecular Ecology* 15, pp. 1797-1806.

Li, T., Chen, X.-L. and Hong, X.-Y. (2009) 'Population genetic structure of *Tetranychus urticae* and its sibling species *Tetranychus cinnabaribus* (Acari: Tetranychidae) in China as inferred from microsatellite data', *Annals of the Entomological Society of America* 102, pp. 674-683.

Li, Y.C., Korol A.B., Fahima T., Beiles A. and Nevo E. (2002) 'Microsatellites: Genomic distribution, putative functions and mutational mechanisms: A review', *Molecular Ecology* 11, pp. 2453-2465.

Linke-Gamenick, I., Vismann, B. and Forbes, V.E. (2000) 'Effects of fluoranthene and ambient oxygen levels on survival and metabolism in three sibling species of *Capitella* (polychaeta)', *Marine Ecology Progress Series* 194, pp. 169–177.

Lister, L.J., Svendsen, C., Wright, J., Hooper, H.L. and Spurgeon, D.J. (2011) 'Modelling the joint effects of a metal and a pesticide on reproduction and toxicokinetics in Lumbricid earthworms', *Environment International* 37, pp. 663-670.

Lopes, I., Baird, D.J., and Ribeiro, R. (2006) 'Genetic adaptation to metal stress by natural populations of *Daphnia longispina*', *Ecotoxcology and Environmental Safety* 63, pp. 275-285.

Lowe, C.N. and Butt, K.R. (2008) '*Allolobophora chlorotica* (Savigny, 1826): evidence for classification as two separate species', *Pedobiologia* 52, pp. 81–84.

Lukkari, T., Taavitsainen, M., Väisänen, A. and Haimi, J. (2004) 'Effects of heavy metals on earthworms along contamination gradients in organic rich soils', *Ecotoxicology and Environmental Safety* 59, pp. 340–348.

Luo, X., Ke, C.-H., You, W.-W., Wang, D.-X. and Chen, F. (2010) 'Molecular identification of interspecific hybrids between *Haliotis discus hannai* Ino and *Haliotis gigantean* Gmelin using amplified fragment-length polymorphism and microsatellite markers', *Aquaculture Research* 41, pp. 1827-1834.

Makova, K.D., Nekrutenko, A. and Baker, R. (2000) 'Evolution of microsatelite alleles in four species of mice (genus *Apodemus*)', *Journal of Molecular Evolution* 51, pp. 166–172.

Mallet, J. (2006) 'Species concepts', in Fox, C.W. and Wolf, J.B. (eds.) *Evolutionary Genetics: Concepts and Case Studies*. Oxford: Oxford University Press, pp. 367-373.

Marinissen, J.C.Y. and Van den Bosch, F. (1992) 'Colonization of new habitats by earthworms', *Oecologia* 91, 371–376.

McMillan A., Bagley M., Jackson S. and Nacci D. (2006) 'Genetic diversity and structure of an estuarine fish (*Fundulus heteroclitus*) indigenous to sites associated with a highly contaminated urban harbor', *Ecotoxicology* 15, pp. 539–548.

McNeilly, T.S. (1968) 'Evolution in closely adjacent plant populations. III. *Agrostis tenuis* on a small copper mine', *Heredity* 23, pp. 99-108.

Mengual, X., Ståhls, G. and Santos, R. (2008) 'First phylogeny of predatory flower flies (Diptera, Syrphidae, Syrphinae) using mitochondrial COI and nuclear 28S rRNA genes: conflict and congruence with the current tribal classification', *Cladistics* 24, pp. 543–562.

Meyers, L.A. and Bull J.J. (2002) 'Fighting change with change: adaptive variation in an uncertain world', *Trends in Ecology and Evolution* 17, pp. 551–557.

Mhatre, G.N. and Pankhurst C.E. (1997) 'Bioindicators to detect contamination of soils with special reference to heavy metals', in Pankhurst, C.E., Doube, B.M. and Gupta, V.V. (eds.), *Biological indicators of soil health*. Wallingford: CAB International, pp. 349–369.

Michl, T., Huck, S., Schmitt, T., Liebrich, A., Haase, P. and Büdel, B. (2010) 'The molecular population structure of the tall forb *Cicerbita alpina* (Asteraceae) supports the idea of cryptic glacial refugia in central Europe', *Botanical Journal of the Linnean Society* 164, pp. 142-154.

Molbo, D., Machado, C.A., Sevenster, J.G., Keller, L. and Herre, E.A. (2003) 'Cryptic species of fig-pollinating wasps: implica- tions for the evolution of the fig-wasp mutualism, sex allocation, and precision of adaptation', *Proceedings of the National Academy of Sciences of the United States of America* 100, pp. 5867–5872.

Morgan, A.J., Kille, P. and Sturzenbaum, S.R. (2007) 'Microevolution and ecotoxicology of metals in invertebrates', *Environmental Science and Technology* 41, pp. 1085–1096.

Morgan, A.J., Stürzenbaum, S.R., Winters, C., Grime, G.W., Aziz, N.A. and Kille, P. (2004) 'Differential metallothionein expression in earthworm (*Lumbricus rubellus*) tissues', *Ecotoxicology and Environmental Safety* 57, pp. 11-19.

Morgan, J.E. and Morgan, A.J. (1988) 'Earthworms as biological monitors of cadmium, copper, lead and zinc in metalliferous soils', *Environmental Pollution* 54, pp. 123-138.

Morgan, J.E. and Morgan, A.J. (1991) 'Differences in the accumulated metal concentrations in two epigeic earthworm species (*Lumbricus rubellus* and *Dendrodrillus rubidus*) living in contaminated soils', *Bulletin of Environmental Contamination and Toxicology* 47, pp. 296-301.

Morgan, J.E. and Morgan, A.J. (1998) 'The distribution and intracellular compartmentation of metals in the endogeic earthworm *Aporrectodea caliginosa* sampled from an unpolluted and a metal-contaminated site', *Environmental Pollution* 99, pp. 167–175.

Moriarty, F. (1983) *Ecotoxicology: the study of pollutants in ecosystems*. New York: Academic Press.

Moritz, C. and Cicero, C. (2004) 'DNA Barcoding: promise and pitfalls', *PLoS Biology* 2, e354.

Muller, L.A.H., Vangronsveld, J. and Colpaert, J.V. (2007) 'Genetic structure of Suillus luteus populations in heavy metal polluted and nonpolluted habitats', *Molecular Ecology* 16, pp. 4728-4737.

Nadeau, D., Corneau, S., Plante, I., Morrow, G. and Tanguay, R.M. (2001) 'Evaluation for Hsp70 as a biomarker of effect of pollutants on the earthworm *Lumbricus terrestris*', *Cell Stress Chaperones* 6, pp. 153–163.

Nahmani, J., Lavelle, P., Lapied, E. and Van Oort, F. (2003) 'Effects of heavy metal soil pollution on earthworm communities in the north of France', *Pedobiologia* 47, pp. 663-669.

Nannoni, F., Protano, G. And Riccobono, F. (2011) 'Uptake and bioaccumulation of heavy elements by two earthworm species from a smelter contaminated area in northern Kosovo', *Soil Biology and Biochemistry* 43, pp. 2359-2367.

Nauta, M. J., and Weissing, F.J. (1996) 'Constraints on allele size at microsatellite loci: implications for genetic differentiation', *Genetics* 143, pp. 1021–1032.

Nevo, E. (1979) 'Adaptive convergence and divergence of subterranean mammals', *Annual Review of Ecology and Systematics* 10, pp. 269–308.

Nosil, P. and Crespi, B.J. (2004) 'Does gene flow constrain adaptive divergence or vice versa? A test using ecomorphology and sexual isolation in *Timema cristinae* walking-sticks', *Evolution* 58, pp. 102–112.

Nosil, P., Crespi, B. and Sandoval, C.P. (2003) 'Reproductive isolation driven by the combined effects of ecological adaptation and reinforcement', *Proceedings of the Royal Society of London Series B-Biological Sciences* 270, pp. 1911-1918.

Novo, M., Almodovar, A., Fernandez, R.M., Gutierrez, M. and Cosin, D.J.D. (2010a) 'Mate choice of an endogeic earthworm revealed by microsatellite markers', *Pedobiologia* 53, pp. 375-379.

Novo, M., Almodóvar, A., Fernández, R., Trigo, D. and Díaz Cosín, D.J. (2010b) 'Cryptic speciation of hormogastrid earthworms revealed by mitochondrial and nuclear data', *Molecular Phylogenetics and Evolution* 56, pp. 507-512.

Nowak, C., Vogt, C., Oetken, M., Pfenninger, M., Schwenk, K., Oehlmann, J. and Streit B. (2009) 'Genetic erosion in tributyltin exposed experimental Chironomus populations', *Environmental Pollution* 157, pp. 881-886.

Osté, L.A., Dolfing, J., Ma, W.-C. and Lexmond, T.M. (2001) 'Cadmium uptake by earthworms as related to the availability in the soil and the intestine', *Environmental Toxicology and Chemistry* 20, pp. 1785–1791.

Owen, J., Hedley, B.A., Svendsen, C., Wren, J., Jonker, M.J., Hankard, P.K., Lister, L.J., Stürzenbaum, S.R., Morgan, A.J., Spurgeon, D.J., Blaxter, M.L. and Kille, P. (2008) 'Transcriptome profiling of developmental and xenobiotic responses in a keystone soil animal, the oligochaete annelid *Lumbricus rubellus*', *BMC Genomics* 9, 266.

Page, R.D.M. (1996) 'TreeView: An application to display phylogenetic trees on personal computers', *Computer Applications in the Biosciences* 12, pp. 357-358.

Palumbi, S.R. (1996) 'Nucleic acids II: the polymerase chain reaction', in D.M. Hillis, C. Moritz and B.K. Mable (eds.) *Molecular Systematics*. Sunderland, MA: Sinauer Associates, Inc., pp. 205–247.

Panova, M., Makinen, T., Fokin, M., Andre, C. and Johannesson, K. (2008) 'Microsatellite cross-species amplification in the genus *Littorina* and detection of null alleles in *Littorina saxatilis*', *Journal of Molluscan Studies* 74, pp. 111–117.

Paoletti, M.G. (1999) 'The role of earthworms for assessment of sustainability and as bioindicators', *Pedobiologia* 14, pp. 1–27.

Pearson, D. (1995) 'Selecting indicator taxa for the quantitative assessment of biodiversity', *Proceedings of the Royal Society B: Biological Sciences* 345, pp. 75–79.

Peijnenburg, W.J.G.M., Baerselman, R., de Groot, A.C., Jager, T., Posthuma, L., Rens P. and Van Veen, R.PM. (1999) 'Relating environmental availability to bioavailability: soil-type-

dependent metal accumulation in the oligochaete *Eisenia andrei*', *Ecotoxicology and Environmental Safety* 44, pp. 294-310.

Peles, J., Towler, W.I. and Guttman, S.I. (2003) 'Population genetic structure of earthworms (*Lumbricus rubellus*) in soils contaminated by heavy metals', *Ecotoxicology* 12, pp. 379-386.

Pérez-Losada, M., Ricoy, M., Marshall, J.C. and Dominguez, J. (2009) 'Phylogenetic assessment of the earthworm *Aporrectodea caliginosa* species complex (Oligochaeta: Lumbricidae) based on mitochondrial and nuclear DNA sequences', Molecular *Phylogenetics and Evolution* 52, pp. 293-302.

Petren, K., Grant, B.R. and Grant, P.R. (1999) 'A phylogeny of Darwin's finches based on microsatellite DNA length variation', *Proceedings of the Royal Society B: Biological Sciences* 266, pp. 321-329

Piearce, T.G. (1978) 'Gut contents of some lumbricid earthworms', *Pedobiologia* 18, pp. 153-157.

Piearce, T.G. (1983) 'Functional-morphology of lumbricid earthworms, with special reference to locomotion', *Journal of Natural History* 17, pp. 95-111.

Piearce, T.G., Langdon, C.J., Meharg, A.A. and Semple, K.T. (2002) 'Yellow earthworms: distinctive pigmentation associated with arsenic- and copper-tolerance in *Lumbricus rubellus*', *Soil Biology and Biochemistry* 34, pp. 1833–1838.

Pielou, E.C. (1991) *After the Ice Age: The Return of Life to Glaciated North America*. Chicago: University of Chicago Press.

Pizl, V. and Josens, G. (1995) 'Earthworm communities along a gradient of urbanization', *Environmental Pollution* 90, pp. 7–14.

Plath, M., Hauswaldt, S., Moll, K., Tobler, M., Garcia de Léon, F.J., Schlupp, I. and Tiedemann, R. (2007) 'Local adaptation and pronounced genetic differentiation in an extremophile fish, *Poecilia mexicana*, inhabiting a Mexican cave with toxic hydrogen sulfide', *Molecular Ecology* 16, pp. 967-976.

Plath, M., Hermann, B., Schröder, C., Riesch, R., Tobler, M., García de León, F.J., Schlupp, I. and Tiedemann, R. (2010) 'Locally adapted fish populations maintain small-scale genetic differentiation despite perturbation by a catastrophic flood event', *BMC Evolutionary Biology* 10, 256.

Pollard, S.J.T., Lythgo, M. and Duarte-Davidson, R. (2001) 'The extent of contaminated land problems and the scientific response', *Issues in Environmental Science and Technology* 16, pp. 1-19.

Pook, C., Lewis, C. and Galloway, T. (2009) 'The metabolic and fitness costs of resistance to copper and zinc toxicity in *Nereis diversicolor*', *Marine Pollution Bulletin* 58, pp. 1063-1071.

Pop, A.A., Cech, G., Wink, M., Csuzdi, C. and Pop, V.V. (2007) 'Application of 16S, 18S rDNA and COI sequences in the molecular systematics of the earthworm family Lumbricidae (Annelida, Oligochaeta)', *European Journal of Soil Biology* 43, pp. S43-S52.

Pop, A.A., Wink, M. and Pop, V.V. (2003) 'Use of 18S, 16S rDNA and cytochrome c oxidase sequences in earthworm taxonomy (Oligochaeta, Lumbricidae)', *Pedobiologia* 47, pp. 428-433.

Posada, D. (2008) 'jModelTest: Phylogenetic Model Averaging', *Molecular Biology and Evolution* 25, pp. 1253-1256.

Posthuma, L., Hogervorst, R. and Van Straalen, N.M. (1992) 'Adaptation to soil pollution by cadmium excretion in natural populations of *Orchesella cincta* (L.) (Collembola)', *Archives of Environmental Contamination and Toxicology* 22, pp. 146–156.

Posthuma, L. and Van Straalen, N.M. (1993) 'Heavy-metal adaptation in terrestrial invertebrates: a review of occurrence, genetics, physiology and ecological consequences', Comparative Biochemistry and Physiology 106C, pp. 11-38.

Postma, J.F. and Groenendijk, D. (1999) 'Adaptation to metals in the midge Chironomus riparius: a case study in the River Dommel', in V.E. Forbes (ed) *Genetics and Ecotoxicology*. Philadelphia, PA: Taylor & Francis, pp. 79-101.

Power, M. and McCarty, L.S. (1997) 'Fallacies in ecological risk assessment practices', *Environmental Management* 31, pp. 370A-375A.

Poynton, H.C. and Vulpe, C.D. (2009) 'Ecotoxicogenomics: emerging technologies for emerging contaminants', *Journal of the American Water Resources Association* 45, pp. 83–96.

Pritchard, J.K., Stephens, M. and Donnelly, P. (2000) 'Inference of population structure using multilocus genotype data', *Genetics* 155, pp. 945–959.

Quirós, L., Piña, B., Solé, M., Blasco, J., López, M.A., Riva, M.C., Barceló, D. and Raldúa, D. (2007) 'Environmental monitoring by gene expression biomarkers in *Barbus graellsii*: laboratory and field studies', *Chemosphere* 67, pp. 1144–1154.

Racey P., Barratt E.M., Burland T.M., Deaville R., Gotelli D., Jones G. and Piertney S. (2007) 'Microsatellite DNA polymorphism confirms reproductive isolation and reveals differences in population genetic structure of cryptic pipistrelle bat species', *Biological Journal of the Linnean Society* 90, pp. 539–550.

Raymond, M. and Rousset, F. (1995) 'GENEPOP 3.4, population genetics software for exact tests and ecumenicism', *Journal of Heredity* 86, pp. 248–249.

Reeves, W.K. (1999) 'New records of cave-dwelling earthworms (Oligochaeta: Lumbricidae, Megascolecidae and Naididae) and other annelids (Aeolosomatida, Branchiobdellida and Hirudinea) in the southeastern United States, with notes on their ecology' *Megadrilogica* 7, pp.65-71.

Reinecke, S.A., Prinsloo, M.W. and Reinecke, A.J. (1999) 'Resistance of *Eisenia fetida* (Oligochaeta) to cadmium after long-term exposure', *Ecotoxicology and Environmental Safety* 42, pp. 75–80.

Reinecke, S.A. and Reinecke, A.J. (1997) 'The influence of lead and manganese on spermatozoa of *Eisenia fetida* (Oligochaeta)', *Soil Biology and Biochemistry* 29, pp. 737-742.

Rico, C., Rico, I. and Hewitt, G. (1996) '470 million years of conservation of microsatellite loci among fish species', *Proceedings of the Royal Society of London, Series B* 263, pp. 549–557.

Rocha-Olivares, A., Fleeger, J.W., and Foltz, D.W. (2001) 'Decoupling of molecular and morphological evolution in deep lineages of a meiobenthic harpacticoid copepod', *Molecular Biology and Evolution* 18, pp. 1088-1102.

Roelofs, D., Morgan, A.J. and Stürzenbaum, S.R. (2010) 'The significance of genome-wide transcriptional regulation in the evolution of stress tolerance', *Evolutionary Ecology* 24, pp. 527-539.

Ronquist, F. and Huelsenbeck, J.P. (2003) 'MrBayes 3: Bayesian phylogenetic inference under mixed models', *Bioinformatics* 19, pp. 1572-1574.

Roy, V., Demanche, C., Livet, A. and Harry, M. (2006) 'Genetic differentiation in the soil-feeding termite *Cubitermes* sp. *affinis subarquatus*: occurrence of cryptic species revealed by nuclear and mitochondrial markers', *BMC Evolutionary Biology* 6, 102.

Rozen, A. (2006) 'Effect of cadmium on life-history parameters in *Dendrobaena octaedra* (Lumbricidae: Oligochaeta) populations originating from forests differently polluted with heavy metals', *Soil Biology and Biochemistry* 38, pp. 489–503

Rozen, S. and Skaletsky, H.J. (2000) 'Primer3 on the WWW for general users and for biologist programmers', in Krawetz, S. and Misener, S. (eds.) *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Totowa, NJ: Humana Press, pp. 365-386.

Satchell, J.E. (1955) 'Some aspects of earthworm ecology', in Kevan, D.K.McE. (ed.) *Soil Zoology: Proceedings of the University of Nottingham Second Easter School in Agricultural Science*. London: Butterworths, pp. 180–201.

Satchell, J.E. (1967) 'Lumbricidae', in Burges, A. and Raw, F. (eds.) *Soil Biology*. London: Academic Press, pp. 259–322.

Savolainen, V., Cowan, R.S., Vogler, A.P., Roderick, G.K. and Lane, R. (2005) 'Towards writing the encyclopaedia of life: an introduction to DNA barcoding', *Proceedings of the Royal Society of London Series B- Biological Sciences* 360, pp. 1805–1811.

Schlichting, C.D. and Smith H. (2002) 'Phenotypic plasticity: linking molecular mechanisms with evolutionary outcomes', *Evolutionary Ecology* 16, pp. 189–211.

Schlick-Steiner, B.C., Seifert, B., Stauffer, C., Christian, E., Crozier, R.H. and Steiner F.M. (2007) 'Without morphology, cryptic species stay in taxonomic crypsis following discovery', *Evolution* 22, pp. 391-392.

Schlick-Steiner, B.C., Steiner, F.M., Moder, K., Seifert, B., Sanetra, M., Dryerson, E., Stauffer, C. and Christian, E. (2006) 'A multidisciplinary approach reveals cryptic diversity in Western Paleartic *Tetramorium* ants (Hymenoptera: Formicidae)' *Molecular Phylogenetics and Evolution* 40, pp. 259–273.

Schlötterer, C., Amos, W. and Tautz, D. (1991) 'Conservation of polymorphic simple sequence loci in cetacean species', *Nature* 354, pp. 63–65.

Schlötterer, C. and Wiehe., T. (1999) 'Microsatellites, a neutral marker to infer selective sweeps', in Goldstein, D.B. and Schlötterer, C. (eds.) *Microsatellites: Evolution and applications*. New York: Oxford University Press, pp. 238-248.

Schon, N.L., Mackay, A.D., Gray, R.A. and Minor, M.A. (2011) 'Earthworms in New Zealand sheep- and dairy-grazed pastures with focus on anecic *Aporrectodea longa*', Pedobiologia 54S, S131-S137.

Schwert, D.P. (1980) 'Active and passive dispersal of lumbricid earthworms', in Dindale, D.L. (ed.) *Soil Biology as Related to Land Use Practices*. Washington DC: E.P.A. pp. 182-189.

Scott-Fordsmand, J.J. and Weeks, J.M. (2000) 'Biomarkers in earthworms', *Reviews of Environmental Contamination & Toxicology* 165, pp. 117–159.

Scourse, J.D., Haapaniemi, A.L., Colmenero-Hidalgo, E., Peck, V.L., Hall, I.R., Austin, W.E.N., Knutz, P.C. and Zahn, R. (2009) 'Growth, dynamics and deglaciation of the last British-Irish ice sheet: the deep-sea ice-rafted detritus record', *Quaternary Science Reviews* 28, pp. 3066-3084.

Sejrup, H.P., Hjelstuen, B.O., Dahlgren, K.I.T., Haflidason, H., Kuijpers, A., Nygard, A., Praeg, D., Stoker, M.S. and Vorren, T.O. (2005) 'Pleistocene glacial history of the NW European continental margin', *Marine and Petroleum Geology* 22, pp. 1111-1129.

Shepeleva, O.A., Kodolova, O.P., Zhukovskaya, E.A. and Striganova, B.R. (2008) 'Genetic diversity of populations of the earthworm *Lumbricus rubellus* (Hoffm.) (Oligochaeta, Lumbricidae)', *Biology Bulletin* 35, pp. 170–177.

Simões, P., Pascual, M., Santos, J., Rose, M.R. and Matos M. (2008) 'Evolutionary dynamics of molecular markers during local adaptation: a case study in *Drosophila subobscura*', *BMC Evolutionary Biology* 8, 66.

Simonsen, V. and Klok, T.C. (2010) 'Genetic and ecological impacts of heavy metal and flooding stress on the earthworm *Lumbricus rubellus* in floodplains of the Rhine river', *Soil Biology and Biochemistry* 42, pp. 270–275.

Sims, R.W. and Gerard, B.M. (1999) *Earthworms: notes for the identification of British species*. Shrewsbury: Linnean Society of London and Estuarine and Coastal Sciences Association.

Sites, J.W. and Crandall, K., (1997) 'Testing species boundaries in biodiversity studies', *Conservation Biology*, 11, pp. 1289-1297.

Slatkin, M. (1987) 'Gene flow and the geographic structure of natural populations', *Science* 236, pp. 787-792.

Somers, C.M., Neudorf, K, Jones, K.L. and Lance, S.L. (2011) 'Novel microsatellite loci for the compost earthworm *Eisenia fetida*: a genetic comparison of three North American Vermiculture Stocks', *Pedobiologia* 54, pp. 111-117.

Sorour, J. and Larink, O. (2001) 'Toxic effects of benomyl on the ultra structure during spermatogenesis of the earthworm *Eisenia fetida*', *Ecotoxicology and Environmental Safety* 50, pp. 180-188.

Spurgeon, D.J. and Hopkin, S.P. (2000) 'The development of genetically inherited Resistance to zinc in laboratory selected generations of the earthworms *Eisenia fetida*', *Environmental Pollution* 109, pp. 193-201.

Spurgeon, D.J., Morgan, A.J. and Kille, P. (2008) 'Current research in soil invertebrate ecotoxicogenomics', in Hogstrand, C. and Kille, P. (eds.) *Comparative toxicogenomics, 2*. London: Elsevier Applied Science (Advances in Experimental Biology), pp. 133-163.

Spurgeon, D.J., Ricketts, H., Svendsen, C., Morgan, A.J. and Kille, P. (2005) 'Hierarchical responses of soil invertebrates (earthworms) to toxic metals stress', *Environmental Science and Technology* 39, pp. 5327–5334.

Spurgeon, D.J., Sturzenbaum, S.R., Svendsen, C., Hankard, P.K., Morgan, A.J., Weeks, J.M. and Kille, P. (2004) 'Toxicological, cellular and gene expression responses in earthworms exposed to copper and cadmium', *Comparative Biochemistry and Physiology C- Toxicology & Pharmacology* 138, pp. 11-21.

Spurgeon, D.J., Svendsen, C., Kille, P., Morgan, A.J. and Weeks, J.M. (2004) 'Responses of earthworms (*Lumbricus rubellus*) to copper and cadmium as determined by measurement of juvenile traits in a specifically designed test system', *Ecotoxicology and Environmental Safety* 57, pp. 54-64.

Spurgeon, D.J., Svendsen, C., Rimmer, V.R., Hopkin, S.P. and Weeks, J.M. (2000) 'Relative sensitivity of life-cycle and biomarker responses in four earthworm species exposed to zinc', *Environmental Toxicology and Chemistry* 19, pp. 1800–1808.

Spurgeon, D.J., Svendsen, C., Weeks, J.M., Hankard, P.K., Stubberud, H.E. and Kammenga, J.E. (2003a) 'Quantifying copper and cadmium impacts on intrinsic rate of population increase in the terrestrial oligochaete *Lumbricus rubellus*', *Environmental Toxicology and Chemistry* 22, pp. 1465–1472.

Spurgeon, D.J., Weeks, J.M. and Van Gestel, C.A.M. (2003b) 'A summary of eleven years progress in earthworm ecotoxicology', *Pedobiologia* 47, pp. 588-606.

Stewart, J.R. and Lister, A.M. (2001) 'Cryptic northern refugia and the origins of the modern biota', *Trends in Ecology and Evolution* 16, pp. 608-613.

Stewart, J.R., Lister, A.M., Barnes, I., and Dalen, L. (2010) 'Refugia revisited: individualistic responses of species in space and time', *Proceedings of the Royal Society of London Series B-Biological Sciences* 277, pp. 661-671.

Streiff, R., Veyrier, R., Audiot, P., Meusnier, S. and Brouat, C. (2005) 'Introgression in natural populations of bioindicators: a case study of *Carabus splendens* and *Carabus punctatoauratus*', *Molecular Ecology* 14, pp. 3775–3786.

Sturmbauer, C., Opadiya, G.B., Niederstatter, H., Riedmann, A. and Dallinger, R. (1999) 'Mitochondrial DNA reveals cryptic oligochaete species differing in cadmium resistance', *Molecular Biology and Evolution* 16, pp. 967-974.

Stürzenbaum, S. R., Parkinson, J., Blaxter, M., Morgan, A. J., Kille, P. and Georgiev, O. (2003) 'The earthworm expressed sequence tag project', *Pedobiologia* 47, pp. 447–451.

Stürzenbaum, S.R., Andre, J., Kille, P. and Morgan, A.J. (2009) 'Earthworm genomes, genes and proteins: the (re)discovery of Darwin's worms', *Proceedings of the Royal Society of London Series B-Biological Sciences* 276, pp. 789-797.

Symondson, W.O.C., Glen, D.M., Erickson, M.L., Liddell, J.E. and Langdon, C.J. (2000) 'Do earthworms help to sustain the slug predator Pterostichus melanarius (Coleoptera: Carabidae) within crops? Investigations using monoclonal antibodies', *Molecular Ecology* 9, pp. 1279–1292.

Taberlet, P., Fumagalli, L., Wust-Saucy, A.G. and Cosson, J.F. (1998) 'Comparative phylogeography and postglacial colonization routes in Europe', *Molecular Ecology* 7, pp. 453-464.

Tabor, G.M. and Aguirre, A.A. (2004) 'Ecosystem Health and Sentinel Species: Adding an Ecological Element to the Proverbial 'Canary in the Mineshaft.'' *EcoHealth* 1, pp. 226-228.

Tamura, K., Dudley, J., Nei, M. and Kumar, S. (2007) 'MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0.', *Molecular Biology and Evolution* 24, pp. 1596-1599.

Terhivuo, J. (1988) 'The Finnish Lumbricidae Oligochaete – fauna and its formation', *Annales Zoologici Fennici* 25, pp. 229–247.

Thissen, D., Steinberg, L. and Kuang, D. (2002) 'Quick and easy implementation of the Benjamini-Hochberg procedure for controlling the false positive rate in multiple comparisons', *Journal of Educational and Behavioral Statistics* 27, pp. 77–83.

Tiunov, A.V., Hale, C.M., Holdsworth, A.R. and Vsevolodova-Perel, T.S. (2006) 'Invasion patterns of Lumbricidae into the previously earthworm-free areas of northeastern Europe and the western Great Lakes region of North America', *Biological Invasions* 8, pp. 1223-1234.

Tobler, M., DeWitt, T.J., Schlupp, I., Garcia de Leon, F.J., Herrmann, R., Feulner, P., Tiedemann, R. and Plath, M. (2008) 'Toxic hydrogen sulfide and dark caves: phenotypic and genetic divergence across two abiotic environmental gradients in *Poecilia mexicana*', *Evolution* 62, pp. 2643–2649.

Turbé, A., De Toni, A., Benito, P., Lavelle, P., Ruiz, N., Van der Putten, W.H., Labouze, E. and Mudgal, S. (2010) *Soil biodiversity: functions, threats and tools for policy makers. Contract 07.0307/2008/517444/ETU/B1. Final report. Bio Intelligence Service, IRD, and NIOO report for the European Commission.* Brussels: European Commission.

Väli, Ü., Dombrovski, V., Treinys, R., Bergmanis, U., Daróczi, S.J., Dravecky, M., Ivanovski, V., Lontkowski, J., Maciorowski, G., Meyburg, B.-U., Mizera, T., Zeitz, R. and Ellegren, H. (2010) 'Widespread hybridization between the Greater Spotted Eagle *Aquila clanga* and the Lesser Spotted Eagle *Aquila pomarina* (Aves: Accipitriformes) in Europe', *Biological Journal of the Linnean Society* 100, pp. 725–736.

Van der Wurff, A.W.G., Isaaks, J.A., Ernsting, G. and Van Straalen, N.M. (2003) 'Population substructures in the soil invertebrate *Orchesella cincta*, as revealed by microsatellite and TE-AFLP markers', *Molecular Ecology* 12, pp. 1349–1359.

Van Oosterhout, C., William, F., Hutchinson, D.P., Wills, M. and Shipley, P. (2004) 'Micro-Checker: software for identifying and correcting genotyping errors in microsatellite data', *Molecular Ecology Notes* 4, pp. 535–538.

Van Straalen, N.M. (2003) 'Ecotoxicology becomes stress ecology', Environmental Science and Technology 37, pp. 324A–330A.

Van Straalen, N.M. and Roelofs, D. (2008) 'Genomics technology for assessing

soil pollution', *Journal of Biology* 7, 19.

Van Straalen, N.M. and Timmermans, M. (2002) 'Genetic variation in toxicant-stressed populations: an evaluation of the "genetic erosion" hypothesis', *Human and Ecological Risk Assessment* 8, pp. 983–1002.

Vandegehuchte, M.B. and Janssen, C.R. (2011) 'Epigenetics and its implications for ecotoxicology', *Ecotoxicology* 20, pp. 607-624.

Vasseur, P. and Cossu-Leguille, C. (2003) 'Biomarkers and community indices as complementary tools for environmental safety', *Environment International* 28, pp. 711-717.

Velavan, T.P., Sahm, S., Schulenburg, H. and Michiels, N.K. (2009) 'High genetic diversity and heterogeneous parasite load in the earthworm *Lumbricus terrestris* on a German meadow', *Soil Biology and Biochemistry* 41, pp. 1591–1595.

Velavan, T.P., Schulenburg, H. and Michiels, N. (2007) 'Development and characterization of novel microsatellite markers for the common earthworm (*Lumbricus terrestris* L)', *Molecular Ecology Notes* 7, pp. 1060–1062.

Vialatte, A., Guiller, A., Bellido, A. and Madec, L. (2008) 'Phylogeography and historical demography of the Lusitanian snail *Elona quimperiana* reveal survival in unexpected separate glacial refugia', *BMC Evolutionary Biology* 8, 339.

Vijver, M.G., Vink, J.P.M., Miermans, C.J.H. and van Gestel, C.A.M. (2003) 'Oral sealing using glue: a new method to distinguish between intestinal and dermal uptake of metals in earthworms', *Soil Biology and Biochemistry* 35, pp. 125-132.

Voua Otomoa, P., Owojori, O.J., Reinecke, S.A., Daniels, S. and Reinecke, A.J. (2011) 'Using estimates of metal bioavailability in the soil and genetic variation of allozymes to investigate heavy metal tolerance in the earthworm *Eisenia fetida* (Oligochaeta)', *Ecotoxicology and Environmental Safety* 74, pp. 2070-2074.

Voua Otomo, P. and Reinecke, S.A. (2010) 'Increased cytotoxic and genotoxic tolerance of *Eisenia fetida* (Oligochaeta) to cadmium after long-term exposure', *Ecotoxicology* 19, pp. 362-368.

Waits, E.R., Bagley, M.J., Blum, M.J., McCormick, F.H. and Lazorchak, J.M. (2008) 'Source-sink dynamics sustain central stonerollers (*Campostoma anomalum*) in a heavily urbanized watershed', *Freshwater Biology* 53, pp. 2061-2075.

Walker, C.H., Hopkin, S.P., Sibly, R.M. and Peakall, D.B. (1996) *Principles of Ecotoxicology*. London: Taylor & Francis.

Wares, J.P., Daley, S., Wetzer, R. and Toonen, R.J. (2007) 'An evaluation of cryptic lineages of *Idotea balthica* (Isopoda: Idoteidae): Morphology and microsatellites', *Journal of Crustacean Biology* 27, pp. 643-648.

Weir, B.S. and Cockerham, C.C. (1984) 'Estimating F-statistics for the analysis of population structure', *Evolution* 38, pp. 1358–1370.

Whalen J.K. (2004) 'Spatial and temporal distribution of earthworm patches in cornfield, hayfield and forest systems of southwestern Quebec, Canada', *Applied Soil Ecology* 27, pp. 143-151.

Whitehead, A., Anderson, S.L., Kuivila, K.M., Roach, J.L. and May, B. (2003) 'Genetic variation among interconnected populations of *Catostomus occidentalis*: implications for

distinguishing impacts of contaminants from biogeographical structuring', *Molecular Ecology* 12, pp. 2817–2833.

Whiteley, A.R., Gende, S.M., Gharrett, A.J. and Tallmon, D.A. (2009) 'Background matching and color-change plasticity in colonizing freshwater sculpin populations following rapid deglaciation', *Evolution* 63, pp. 1519–1529.

Whitman, D.W. and Agrawal, A.A. (2011) 'What is phenotypic plasticity and why is it important?', in Whitman, D.W. and Ananthakrishna, T.N. (eds.) *Phenotypic plasticity of insects: Mechanisms and consequences*. Enfield, NH: Science Publishers, Inc., pp. 1-63.

Williams, H.C., Ormerod, S.J. and Bruford, M.W. (2006) 'Molecular systematics and phylogeography of the cryptic species complex *Baetis rhodani* (Ephemeroptera, Baetidae)', *Molecular Phylogenetics and Evolution* 40 pp. 370–382.

Wilson, J.A., Lochmiller, R.L. and Janz, D.M. (2006) 'Population dynamics of cotton rats (*Sigmodon hispidus*) inhabiting abandoned petroleum landfarms in Oklahoma, USA', *Ecotoxicology* 15, pp. 19-30.

Xie, L. and Klerks, P.L. (2004) 'Changes in cadmium accumulation as a mechanism for cadmium resistance in the least killifish *Heterandria formosa*', *Aquatic Toxicology* 66, pp. 73-81.

Zaitsev, A.S. and van Straalen, N.M. (2001) 'Species diversity and metal accumulation in oribatid mites (Acari, Oribatida) of forests affected by a metallurgical plant', *Pedobiologia* 45, pp. 467–479.

Zardoya, R., Vollmer, D., Craddock, C., Streelman, J., Karl, S. and Meyer, A. (1996) 'Evolutionary conservation of microsatellite flanking regions and their use in resolving the phylogeny of cichlid fishes (Pisces: Perciformes)', *Proceedings of the Royal Society of London Series B-Biological Sciences* 263, pp. 1589–1598.

Zeng, Y.-F., Liao, W.J., Petit, R.J. and Zhang, D.-Y. (2010) 'Exploring species limits in two closely related Chinese oaks', *PLoS ONE* 5, e15529.

Appendix A- Site locations

Site	Site code	Chapter	Latitude	Longitude	Site description
Gloversfield	GF	2	51:18:42N	02:47:33W	Rough pastureland near abandoned Zn mine
Hallen Hill	HH	2	51:30:57N	02:39:28W	Decidous woodland near abandoned Pb/Zn smelter
Ystwyth Source	YS	2	52:21:57N	03:41:44W	Stream bank
East Cottage	EC	2	52:21:35N	03:44:59W	Derelict building at abandoned Pb/Zn mine
Wemyss	WM	2	52:20:58N	03:53:14W	Stream bank near abandoned Pb/Zn mine
Clydach 1	CL1	4	51:41:47N	03:53:18W	Decidous woodland and rough grassland close to Ni smelter
Clydach 2	CL2	4	51:42:30N	03:51:52W	Short grassed riverbank
Clydach 3	CL3	3 and 4	51:42:25N	03:50:48W	Rough pastureland
Clydach 4	CL4	4	51:42:15N	03:51:34W	Horse paddock with little grass
Clydach 5	CL5	4	51:41:47N	03:53:20W	Derelict garden across road from Clydach Ni smelter
Clydach 8	CL8	2	51:41:48N	03:52:49W	Playing field
Clydach 11	CL11	4	51:43:54N	03:49:59W	Canal bank
Clydach 12	CL12	4	51:41:59N	03:53:11W	Small mixed woodland
Rudry 1	RU1	4	51:34:24N	03:10:60W	Peaty rough grass
Rudry 2	RU2	4	51:34:29N	03:11:04W	Peaty rough grass
Rudry 3	RU3	3 and 4	51:34:19N	03:10:53W	Rough grass
Rudry 5	RU5	4	51:33:48N	03:10:47W	Rough pasture
Rudry 6	RU6	4	51:33:37N	03:11:17W	Short cattle-grazed grassland
Rudry 8	RU8	4	51:34:08N	03:10:39W	Stream bank under mixed woodland
Dinas Powys 1	DP1	4	51:26:40N	03:13:33W	Dried stream bank under mixed woodland
Dinas Powys 2	DP2	4	51:27:02N	03:13:44W	Rough grass near river bank under decidous woodland
Dinas Powys 3	DP3	4	51:26:18N	03:13:12W	Edge of grassy field under decidoud woodland
Dinas Powys 6/ Pontcanna	DP6/PT	2 and 4	51:29:37N	03:12:18W	Mound of topsoil leaves on mowed park land

Site	Site vegetation	Earthworm species diversity
Gloversfield	-	-
Hallen Hill	-	-
Ystwyth Source	-	-
East Cottage	-	-
Wemyss	-	-
Clydach 1	Birch and ash trees	*D. rubidus* and some *L. festivus*
Clydach 2	-	Highly diverse including *A. calignosa* and *A. chlorotica*
Clydach 3	Brambles and *Juncus* in grass, hazel trees	Few other species
Clydach 4	-	Some *L. festivus*
Clydach 5	-	Low earthworm numbers, few other species
Clydach 8	-	Highly diverse, many *L. festivus*
Clydach 11	-	High abundance of *L. festivus*
Clydach 12	-	-
Rudry 1	Bilberry	-
Rudry 2	Bilberry, braken and brambles	-
Rudry 3	Bracken and brambles	-
Rudry 5	Ash trees	Highly diverse
Rudry 6	-	Some *L. festivus*
Rudry 8	Buttercups in grass, alder trees	-
Dinas Powys 1	Hazel trees	Some *E. tetraedra*
Dinas Powys 2	Alder trees	-
Dinas Powys 3	Ash trees	Some endogeic species
Dinas Powys 6/ Pontcanna	-	Highly diverse (Several *Lumbricus* species, *S. mammalis*, *D. rubidus*)

Site	Soil pH	Soil Ca (mg/kg)	Soil Ni (mg/kg)	Soil Cd (mg/kg)	Soil Pb (mg/kg)	Soil Zn (mg/kg)	Soil Cu (mg/kg)
Gloversfield	6	-	-	350	6340	43000	50
Hallen Hill	-	-	-	23	1432	1010	-
Ystwyth Source	6	-	-	3	1442	790	35
East Cottage	7	-	-	49	6043	7556	-
Wemyss	7	-	-	32	9081	17376	-
Clydach 1	-	1427	4071	0	333	122	849
Clydach 2	-	8646	80	1	158	485	106
Clydach 3	-	677	70	1	87	102	67
Clydach 4	-	2064	109	2	89	173	83
Clydach 5	-	5275	2191	2	486	683	1528
Clydach 8	-	1133	319	1	133	142	170
Clydach 11	-	99366	34	1	68	201	44
Clydach 12	-	-	-	-	-	-	-
Rudry 1	-	281	9	0	147	52	24
Rudry 2	-	293	6	0	88	41	13
Rudry 3	-	536	14	0	87	55	25
Rudry 5	-	25537	21	1	100	155	20
Rudry 6	-	6366	24	2	225	377	22
Rudry 8	-	2759	17	2	204	289	20
Dinas Powys 1	-	6137	17	0	53	213	20
Dinas Powys 2	-	9738	21	1	77	188	26
Dinas Powys 3	-	6356	24	1	101	200	25
Dinas Powys 6/ Pontcanna	-	28674	16	1	95	174	22